The Alberta Journal of Educational Research

Vol. X, No. 2 *June, 1964*

THE COMMITTEE ON EDUCATIONAL RESEARCH
Faculty of Education
University of Alberta

AN OBJECTIVE COMPARISON OF ACHIEVEMENT IN THE BASIC SUBJECTS FOR MATCHED GROUPS OF CHILDREN IN MANCHESTER, ENGLAND AND EDMONTON, ALBERTA

J. A. YOUNG
AND
M. D. JENKINSON
University of Alberta

The continuing criticism, which belittles education on this continent in comparison with the European, demands that objective evidence concerning achievement in the basic subjects should be obtained. Comparative studies of achievement not only serve as a firm basis for judgment, but also help to illumiate the effects of differing school organization, methods and materials upon achievement.

The Purpose

The purpose of this study was to compare achievement in the basic subjects (reading, arithmetic and language) for matched groups of children in Manchester, England and Edmonton, Alberta. Comparisons were also made on the basis of sex, intellectual superiority and the specific strengths and weaknesses in achievement of two age groups in the respective cities.

Related Research

No previous comparative study of the educational achievements of selected groups of children from the two nations compared in this study was located in the literature. However, a number of similar studies have compared European educational achievement particularly British with those in other geographical regions of this continent.

Scholl (4) compared the reading achievement of children of various ages in Burton-upon-Trent, Staffordshire, England, with the age norms of the Stanford Achievement Tests, which have been standardized on American children. No significant differences in the total test scores for the two groups were identified.

Preston (3) compared the reading achievement of fourth and sixth grade pupils in Wiesbaden, Germany and Philadelphia, U.S.A. The results of his study cast doubt on the validity of unqualified statements that German children read better than American children, or that the incidence of reading retardation is greater in the United States than in Germany.

Studies to compare the achievement in arithmetic of selected groups of 11-year old children in the United States and England have been completed by Buswell (1) and Tracy (5). The tests used (numerical items and word problems) produced significantly superior results for the children in the English samples. Similar results were obtained by Kramer (2) in his comparative study of children at various age levels of the elementary school.

The twelve nation study by U.N.E.S.C.O. was published in the same month in which this study was completed.

All comparative research studies that have been completed represent only a beginning in this rapidly expanding field.

Collection of Data

The Sample

The subjects for his study were 169 nine-year old children and 173 fourteen-year old children in each city. The schools (four in each city) were selected on the recommendations of education authorities in both cities as being representative of the total school population of the chosen age groups in each city. Comparability of socio-economic status was also considered in the selection of the samples. The groups were also matched on the basis of non-verbal intelligence.

The Testing Instruments

In order to reduce the possibility that children in either city would be more familiar with the tests, both English and American tests were used. Tests used, that have been published and standardized in England were the *Schonell Graded Word Spelling Test*, the *Watts Sentence Reading Test* and the *Vernon Sentence Reading Test.* Tests used that have been published and standardized in the United States were the *California Short-Form Test of Mental Maturity,* and the complete battery of the *California Achievement Tests* at the elementary and junior high school levels.

Treatment of Data

Mean scores and variances of the total test scores and sub-test scores were calculated and compared for both age groups in both cities. Similar calculations were made and compared for boys and girls separately at both age levels in both cities, and for those children in both cities whose I.Q. scores on the *California Short-Form Test of Mental Maturity* exceeded 130.

The significance of the differences between the mean scores for each of the groups was tested by the conventional "t" test of significance at the accepted one per cent and five per cent levels.

TABLE I

SUMMARY OF THE TESTS OF SIGNIFICANCE FOR THE DIFFERENCES IN THE MEAN SCORES AND VARIANCES FOR THE TOTAL GROUP AND THE SUB-GROUPS IN THE TWO CITY SAMPLES AT BOTH AGE LEVELS FOR ALL TESTS ADMINISTERED

	C.A.	Lang. I.Q.	Non-Lang. I.Q.	Total I.Q.	Watts	Schonell	Rdg. Vocab.	Rdg. Comp.	Total Rdg.	Arith. Reas.	Arith. Fund.	Total Arith.	Mech. of Eng.	Spelling	Total Lang.	Total Battery
Nine-Year Olds:																
Means																
Total Group Edmonton--Manchester	E .01	N.S.D.	N.S.D.	N.S.D.	N.S.D.	E .01	E .01	E .01	E .01	E .01	M .01	N.S.D.	E .01	E .01	E .01	E .01
Boys $E_9 - M_9$	E .05	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	E .05	E .01	E .01	E .05	N.S.D.	N.S.D.	E .05	N.S.D.	E .01	E .01
Girls $E_9 - M_9$	E .01	N.S.D.	N.S.D.	N.S.D.	N.S.D.	E .01	E .01	E .01	E .01	E .01	M .01	N.S.D.	E .01	E .01	E .01	E .01
Variances																
Total Group $E_9 - M_9$	E .02	M .10	N.S.D.	M .02	N.S.D.	M .02	M .02	M .02	M .02	M .02	M .02	M .02	N.S.D.	M .02	M .02	M .02
Boys $E_9 - M_9$	E .02	N.S.D.	M .10	M .02	N.S.D.	M .10	M .02	N.S.D.	N.S.D.	M .02	M .02	M .02	N.S.D.	M .10	M .10	M .02
Girls $E_9 - M_9$	E .02	N.S.D.	N.S.D.	N.S.D.	N.S.D.	M .02	M .02	M .10	M .10	N.S.D.	M .02	M .02	N.S.D.	M .02	M .10	M .02
Medians:																
Gifted (over 130)	N.S.D.	N.S.D.	N.S.D.	N.S.D.	M .10	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	M .01	M .05	E .01	N.S.D.	E .05	N.S.D.
Fourteen-Year Olds:					Vernon											
Means																
Total Group $E_{14} - M_{14}$	N.S.D.	E .01	N.S.D.	N.S.D.	N.S.D.	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01
Boys $E_{14} - M_{14}$	E .01	N.S.D.	N.S.D.	N.S.D.	N.S.D.	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01
Girls $E_{14} - M_{14}$	N.S.D.	E .01	N.S.D.	N.S.D.	N.S.D.	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01	E .01
Gifted (over 130)	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.
Variances																
Total Group $E_{14} - M_{14}$	E .02	M .02	N.S.D.	N.S.D.	M .10	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02
Boys $E_{14} - M_{14}$	E .02	M .10	N.S.D.	N.S.D.	N.S.D.	M .02	M .02	M .10	M .02	M .02	M .10	M .02	M .10	M .02	M .02	M .02
Girls $E_{14} - M_{14}$	E .02	M .10	N.S.D.	N.S.D.	N.S.D.	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02	M .02

M -- Significant at the indicated level in favour of Manchester.
E - Significant at the indicated level in favour of Edmonton.
N.S.D. - No significant difference.

The significance of the differences between the variances of the two groups was calculated and tested by the conventional "F" test of significance at the accepted two per cent and ten per cent levels.

The Results

A summary of the results obtained is presented in the table.

Differences in Mean Scores:

Schonell Spelling Test: Mean scores for the Edmonton children were significantly higher than those of the comparable groups in the Manchester sample. The usual sex differences in spelling achievement with the girls' mean scores significantly exceeding those of the boys were maintained for the groups in both cities.

California Achievement Tests: Mean scores on the separate sub-tests were significantly higher for the Edmonton children than the Manchester children with the exception of the Arithmetic Fundamentals sub-tests.

General significant differences in the mean achievement scores in favor of the Edmonton children for the total groups in the samples from each city were maintained when the sex sub-groups were compared, but tended to disappear when only the achievement of those children in the sample from each city with I.Q. scores above 130 were compared.

Watts and Vernon Reading Tests: No significant differences in the mean scores of the children in the samples from the two cities were identified on those tests.

Differences in the Variances:

The variances of the achievement scores on all tests for the total groups were almost uniformly significantly greater for the Manchester groups than for the Edmonton groups.

Conclusions

Difficulty with the format of the tests might be a contributing factor to the generally lower performance of the Manchester children. The items on the *California Achievement Tests* are of the multiple-choice type, and children in Edmonton schools have considerably greater experience with this type of test item than do children in Manchester schools. However the differences in the mean achievement scores for the two national groups do appear to be real. These differences are highly significant at both age levels, and are even more significant at the fourteen-year old level than at the nine-year old level.

The possibility that these objective tests provide a more valid measurement of the attainments of the more prescriptive and uni-

form Edmonton school curriculum than of the indefinite and varied curriculum of the Manchester schools should be considered. It should also be emphasized that these test results apply only to achievement in the three basic subjects, and might be quite different if general tests of understanding in other content subjects of the school curriculum were administered. Tests of other aspects of achievement in the three basic subjects (e.g. written language, depth and breadth in vocabulary, and the highest levels of reading comprehension) might also give quite different comparative results than those attained in this study.

Though the mean differences of scores were in favor of Edmonton, it is important to note that the differences in the variances which were statistically significant were all in favor of the Manchester groups. This suggests that the English system appears to produce greater variability in the range of achievement, perhaps because of the greater emphasis on individual differences.

Differences in promotion policies in the educational systems of the two cities also favor the children in the Edmonton sample. Promotion is automatic in Manchester schools and all nine-year old children and fourteen-year old children regardless of their achievement level were included in the sample from that city. Since comparable grade levels were chosen for the Edmonton sample, and promotion from grade level to grade level in Edmonton is partly dependent upon achievement, lowest achievers in the Edmonton schools may have been omitted from the sample due to the fact that they had not reached grade levels four and nine, from which the sample was drawn.

The importance of quick, complex mental calculations in arithmetic in English society, resulting in increased emphasis on drill in arithmetic fundamentals in Manchester schools would appear to explain the superiority of the Manchester children on this particular sub-test.

The fact that no significant differences were found in the mean achievement scores of the children on the reading tests that were published and standardized in England might be partly explained by the limited sampling of reading skills tested. Both of these tests provide a sample of the reading ability of the children on a completion type test of reading comprehension only, in a limited time. Also these two sentence reading tests appear to be less closely related to any definite school curriculum, and may therefore provide the best unbiased score for reading ability for both of the two selected national groups.

The superiority of the children in the Edmonton sample at both age levels on the *Schonell Spelling Test* appears to be related to the

marked differences in the spelling programs for the schools in the two cities. A very definite, sequential program in spelling is prescribed for the Edmonton schools with specific exercises and activities in spelling outlined for each day of the week. The Manchester spelling program represents almost a complete opposite, taking an entirely incidental approach to the teaching of spelling. Comparisons of spelling errors made in written language by children in the two cities might be quite different than those attained on this spelling dictation test.

Implications

In general, the results of this study cast doubt on the frequently stated assumption regarding the superiority of British education over any that is given on this continent. However certain implications are suggested.

(1) *Curriculum and Teaching Procedures*:

The organized prescriptive program of instruction in the Edmonton schools, authorized and controlled by the Alberta Department of Education seems to produce a higher average performance rating in the three basic subjects for the majority of students than the much freer, indefinite program of instruction in the Manchester schools. However the flexible curriculum in the Manchester schools seems to widen the range of the pupils' achievement in the basic subjects significantly more than the instructional program provided for the school population of Edmonton.

These results present fundamental general questions to all educators concerned with any aspect of curriculum construction for the schools in either of the two selected countries. Is it possible to provide a school curriculum that will maintain high average performance ratings in the three basic subjects, and at the same time provide many additional opportunities for children to achieve to their maximum capacity in all aspects of the schools' program? Can official direction in matters of curriculum from centralized educational authorities produce uniform higher achievement in the basic subjects without sacrificing the provisions for individual students to proceed through the school program at their own rate?

(2) *Age of School Entry*:

The one year earlier starting age for most Manchester children as compared to the Edmonton starting age does not seem to be an advantage in the acquisition of skills and knowledge in the three basic subjects tested by the chosen battery of tests in this experiment. However, it does not seem fair to dismiss the importance of a planned program of learning experiences for children of chronological age-five years, simply on the basis that such a pro-

gram does not produce significant differences in measured achievement in the basic subjects at nine and fourteen years of age. Recent studies in child development have confirmed the importance of the reading readiness programs. Provisions for many experiences in oral language and for building a background of experience for children from culturally-deprived homes are two of the most essential values from such programs.

However, fundamental educational issues remain to be settled. Would the general learning and development of children throughout the schools' program be improved by definite provision for regular school attendance for all children at five years of age? If so, what types of experiences should be provided to best promote the latter development of the skills, abilities, attitudes and appreciations that are desirable outcomes of the total educational program provided by the schools? Some obvious influences on school achievement in the three basic subjects should accrue from one extra year at school if this extra year were used in the most effective possible manner.

(3) *Streaming*:

The provision for the "streaming" of all children into separate classes, with differentiated programs of instruction according to their achievement and intellectual ability, that is common practice in Manchester schools, appears to widen the range of individual differences in achievement in the three basic subjects, measured in this experiment. The fundamental educational issue that seems apparent from the results of this study is the need for further investigation into the best administrative procedure for providing for real differentiation of the instructional program according to the abilities of the children in the schools without sacrificing a high level of general or uniform achievement. Further investigations into differentiation of the teaching methods and materials used with different groups of children as well as differentiation of the basic curriculum content are needed in both of the selected countries.

(4) *Intellectually-Gifted Children*

The results of this experiment, indicating insufficient challenge to the intellectually-gifted students in the Edmonton sample at both age levels should cause considerable concern to educators on the North American Continent. General questions in this area with which educators in both of the selected countries are faced include the following: What additional provisions should be made in the program to meet the educational needs of intellectually-gifted children? Is is possible to meet these needs within the regular school program provided for all students, or is it necessary to segregate this particular group of children into a completely

separate school organization, in order to provide them with the best type of school program to ensure their maximum educational development?

(5) *Books and Other Teaching Materials*:

Since the results of this study show an almost uniform wider range in the achievement at both age levels of the Manchester children than the Edmonton children, the influence of the use of compulsory authorized textbooks in the three basic subjects on maintaining a greater degree of uniformity for the achievement scores of children in such a prescriptive program should be investigated. Is it possible to maintain a high mean performance in the basic subjects without restricting the core of the curriculum in the basic subjects to definite authorized textbooks? What influence does more official direction in the choice of textbooks have on increasing general achievement in the three basic subjects, while at the same time safeguarding the opportunities for a freer choice of textbooks and teaching aids for children at both of the extreme ends of the children's intellectual ability scale? If the use of prescribed textbooks contributes to the subjection of all Edmonton students to a "lock-step" curriculum, a re-examination of this prescription is very necessary.

The measurement of comparative educational achievement through objective testing has been receiving considerable attention recently. The results of this testing not only provide evidence of comparable levels of achievement but also pose many interesting questions concerning curriculum, methods and materials.

REFERENCES

1. Buswell, G. T. "A Comparison of Achievement in Arithmetic in England and Central California," *The Arithmetic Teacher,* 5:1-9, February, 1958.
2. Kramer, H. "Arithmetic Achievements in Iowa and the Netherlands," *Elementary School Journal,* 59:258-263, February, 1959.
3. Preston, Ralph C. "A Comparative Study of the Reading Achievement of German and American Children," *International Reading Association Conference Proceedings,* 6:109-112, 1961.
4. Scholl, Geraldine T. "The Reading and Spelling Achievement of a Group of English Children," *International Reading Association Conference Proceedings,* 6:107-109, 1961.
5. Tracy, Neal H. "A Comparison of Test Results, (North Carolina, California and England)," *The Arithmetic Teacher,* 6:199-202, October, 1959.

A COMPARATIVE STUDY OF C. P. SNOW'S "TWO CULTURE" CONTROVERSY

Donald F. Swift
and
Thomas E. Linton

Introduction

In recent years, a controversy has developed over what C. P. Snow has termed the "two culture" division in human understanding. He saw this division as a product of over-specialization within the schools and universities. This specialization started very early in England, continued throughout the individual's university training and produced two quite distinct cultures among the educated élite (1).

Professor Snow has argued that this separateness between the "arts" and the "science" culture produces a very dangerous condition in present day society. He contends that the two cultures represent different cognitive and conceptual approaches to human problems. The individual with an arts background is generally unable to understand even the most rudimentary scientific developments. His training has promoted a very limited awareness of modern technical and scientific discoveries.

This lack of understanding tends to be encouraged by the attitudes which an "arts" man learns during his university study. The same may be said of the "science" major in that he frequently knows little about the "arts", and has prejudiced feelings towards people majoring in the arts areas and their disciplines.

These attitudes are promoted by the divisions of interest which are present in various university departments. Not only is there a dichotomy between the "arts" and the "sciences", but there are important differences between disciplines within a particular field. These differences may be emphasized by the students' professor and by the more advanced students in a discipline who informally make sure that the prejudices and partisan values of their field are accepted by the novitiate (2).

The central problem is one of communication between those who have specialized in the Arts and those majoring in the sciences. Both of these individuals may move towards leadership positions in their societies, and be seriously handicapped in their ability to understand the ideas and values of the other individual. At the present time many of the most important decisions affecting the survival or destruction of human life are made by the scientist. The arts man frequently is the administrator who smooths the path for

the scientific act. He seldom understands the technical ramifications of the orders he expedites. His questions about the meaning of a particular move are usually viewed by the scientifically trained staff member as peripheral to the problem itself, if not as essentially unrelated to the issue as perceived by the technician. This further lessens the possibility of good communication between the scientist and the "arts" administrator (3). It also serves to increase the social and political power of the technical specialist, and to lessen the influence of the nonspecialist.

Historically the liberal arts were in the favored position. The cultured and successful individual had only to major in the arts and evidence little interest in the sciences. European tradition still tends to place greatest emphasis on the cultured person as one who knows literature, history and philosophy ,and has an active interest in the fine arts. Technical and scientific training were not traditionally endowed with the same prestigeous aura as were the so-called "Liberalizing" arts (4).

In academic circles it is traditionally assumed that an "arts" background is essential to liberate an individual from his personal and parochial past. However, studies of this assumption have indicated that college training may serve to reinforce old prejudices, and present the student with a more sophisticated means of maintaining his old interests (5).

Method

In an attempt to examine C. P. Snow's hypothesis about the "Two Culture" controversy, Kenneth Richmond in England has developed a Culture Test (6). He administered the test to grammar and public school boys in the sixth form, to students in the university education department, and to students in women's training colleges. He thus provided a series of average scores indicating certain levels of attainment for different student bodies in England. The present authors thought it would be interesting to give the same test in similar educational institutions in Canada and the United States and compare these to the scores achieved in the English schools.

It is difficult to compare educational institutions in different cultures. For the effects of the culture itself as well as the many contrasts in academic approaches make these comparisons artificial. Granting these conditions the authors selected the following institutions:

1. A Provincial University, recently established in one of the larger cities (270,000) in the prairie provinces of Canada.
2. A California "State College of Applied Arts and Science" which has one of the larger teacher training divisions in the states. The college is located in a city of over 3,000,00 people.

3. A major State University located in one of America's larger Eastern seaboard cities (population over 6,000,000).

In general the comparisons were focussed on teachers in training, although additional comparisons were made between teaching majors and non-teaching majors in the Canadian University.

The culture test was composed of two parts with twenty questions in each part. One part was made up of general questions from the arts, while the second part contained more specific questions from the sciences. Mr. Richmond assumed that the mean scores of different groups on the arts and science parts of the test would be an indication of their general level of knowledge in these respective fields.

This article is not presented as a critique of the assumptions of C. P. Snow for it is felt that a different kind of instrument needs to be developed for that purpose. However, it was felt that the Culture Test was indicative of certain features of British secondary education, and in that sense a general comparison with Canadian and American students would be worth presenting.*

The Culture Test was administered to 454 first year students, 47 second year students, and 79 fourth year students in the Canadian Provincial University. In addition 165 students in Education were tested at the California State College, and 111 Education majors attending the American State University.

TABLE I
ENGLAND

	Arts	Science	Total
Public School Sixth Form (17-18 years)	7.3	10	17.3
Grammar School Sixth Form (17-18 years)	5.4	10.9	16.3
Two University Education Departments	7.4	8.6	16
	6.8	9.3	16.1
Three Women's Training Colleges	4.4	5.6	10
	4.7	6.0	10.7
	3.2	3.7	6.9

*Because the Culture Test used in this research was not scientifically constructed and validated, the authors felt that statistical testing for the significance of the differences with the mean scores was not justified. All remarks regarding differences, therefore, are impressionistic.

Results

The results are presented as the average number of questions correctly answered by a particular group of students. The arts and science average scores for each group are reported separately and then as a total out of a possible maximum of forty questions in the test. The tables that follow indicate the scores on the Culture Test obtained in (1) England, (2) Canada, and (3) America. After each of the tables there is brief discussion of the possible implications of the scores.*

Discussion

The Women's Training Colleges all received low scores. These were not university students, the large majority of them are not qualified for university entrance. However, their training does suit them for teaching in the elementary schools. This would be a further indication that academically weaker students frequently enter the teaching profession as elementary school teachers.

We see that the Sixth Form scores are similar to those obtained in a University Department of Education which could mean that the British university students (especially in science) tend not to go into teaching. There seems to be less evidence for this in Canada and the United States. It would be expected that the Public School students would receive higher scores. They tend to come from more successful business and professional families, and are a highly selected group. Their schooling is primarily university committed, and the tone of school life emphasizes the importance of humanistic and scientific learning. The grammar school group represents a lower socio-economic level than the public school group, but there may be more high achieving individuals among these students. For many "average home" students gain access to the grammar schools on the basis of their eleven plus examinations and academic record. This would mean that the grammar school is viewed as a primary ladder of social mobility for the high achieving less affluent student (7).

One would expect the University Education Departments to do better than the Public and Grammar School students. The University students are older and have finished their undergraduate training. They have experienced the many cultural advantages associated with student life. The differences between the University Education Departments and the Public and Grammar School scores may be viewed in these terms. The Public Schools have traditionally been training centers for the children of the élitist

*The material in this study is presented as an interesting note on education in three societies. It does not pretend that these figures may be used as valid reflections on the educational efforts in these countries. A more rigorous sampling design and testing approach would be required if a serious attempt at cross cultural evaluation were to be made.

families in British society. They represent social as well as academic standards not within realistic reach of the average family. It is obvious that the son of an "establishment" family has many cultural advantages not present in the background of children who enter the Grammar Schools, or those who choose to become teachers.

This may account for the difference between the Grammar School arts score and the Public School score. Further the teachers in training in the University achieve about the same level as the Public School students, even though they are older and have had more academic experience. The Public Schools, the Grammar Schools and the University Education students have similar science scores, probably because this involves more straight forward retention of "school taught" material. However, it should be noted that the University students are again behind the Public and Grammar school students. Is this an indication that British teachers in training are derived from a lower achieving academic group?

TABLE II
CANADA

Western Canadian Provincial University	Arts	Science	Total	N
1st year Education	2.11	7.40	9.51	65
1st year Engineering	1.95	8.91	10.90	91
1st year Chemistry	2.07	9.42	11.50	114
1st year English	2.20	7.09	9.29	100
1st year Commerce	1.62	8.62	10.22	84
1st year Average	1.97	7.77	9.74	454
2nd year Education (Jr. E.)	2,64	7.8	10.44	47
B.Ed. Fourth Year	3.87	11.33	15.2	79

*The five different departmental areas making up the first year average are not mutually exclusive categories. Many of the students taking the Chemistry, English and Commerce first year courses may also become teachers. The groups making up the "First Year Average" represents a near total sample of the first year university class (N = 454).

Discussion

The scores of the Canadian first year students do not differ markedly from one department to another. Most significant are the very low scores on the arts questions for all the first year students. One example of this may be seen when a single group is examined. Among the education students there were twelve who scored zero out of the twenty questions, and another thirteen who obtained one out of the twenty for the arts questions. The second year education students showed little improvement, while the fourth year teaching majors evidenced some improvement, but were still very low in the arts area.

What do these low scores on the arts questions represent? The inability to answer more than one or two of the twenty items may indicate a cultural and educational deficiency in the students' background. Even when one grants that the arts questions may be somewhat esoteric, the very low scores are difficult to interpret.

Most of the arts questions are based on information which would be found in the literary and cultural pages of the better magazines and newspapers. Further they would be based on a general level of cultural and literary knowledge which is more characteristic of certain urban centers than of others. In this sense one would expect that the university students in Montreal and Toronto would have higher mean scores on the arts questions than similar students from the universities in the prairie provinces.

It may be argued that a low arts score is not a serious matter for the non-education major. But when one realizes that many of the education majors will enter the teaching profession at the end of their second university year, the situation is worthy of notice. It is only a very recent development that two years are required before commencing to teach in the public schools in this province. Previously one year of preparation for teaching was sufficient, and a large number of teachers in the schools began with this level of training.

When one considers the small amount of education the teachers have experienced, and the "cultural" level of development in the region, the low arts scores are more readily understood. The question that results is what will these teachers be able to pass along to their students in the way of cultural information?

The difference between the scores of the first and fourth year education majors seems to provide some argument in favor of a four year preparation of teachers. However, the small amount of improvement on the arts questions would not indicate much of a return for three more years of university study. The small improvement may indicate something about the "cultural" life in the university apart from the formal instruction in the classroom.

On the positive side the first year students have been well prepared in the "science" area. They have evidently recalled the factual material presented in their high school science classes.

In considering the fourth year B.Ed. performance the same trends are noted. The arts scores are very low for individuals who have been involved in university life for four years.* It may be questioned whether the intensive cramming of material for a large number of courses in seven month sessions is the most beneficial approach which can be offered the intending teacher?

But on the science questions the fourth year Canadian students did better than any of the other schools tested. This is partly a result of the fact that many of these students were science majors. Here the substantial gap between the first year and fourth year promises a valuable increment in scientific knowledge for the student.

TABLE III
AMERICA

	Arts	Science	Total	N
1. California State College				
Education—Undergraduate	5.10	7.50	12.60	72
Education—Graduate Students	6.10	8.81	14.91	93
2. Eastern State University				
Education—Undergraduate	5.25	8.12	13.37	51
Education—Graduate Students	7.07	10.98	18.05	60

Discussion

The two colleges represented in Table III are similar in certain ways. They both cater to an urban student, many of whom work on a full or part time basis in addition to their university studies. A large number of the students at both colleges come from minority groups who attend the low cost state subsidized schools as a means of improving their social and economic position. Further there is a large amount of "cultural" activity in these cities, and it would be difficult for a socially mobile person to avoid even casual contact

*These are generally not students who work apart from their university studies. Their leisure is not committed to economic responsibilities as is the case for many of the students in the urban colleges tested in the U.S.

with cultural information. There is literally an endless variety of entertainment at all cultural levels available without effort to the public. There is an abundance of consumer goods, recreational alternatives, and varieties of living styles within reach of many of these students.

These living conditions are different from those present in the western Canadian University where social and cultural limitations are more obvious. Considering the level of the culture in these American cities the scores on the Arts questions are not very impressive. The graduate student total score at the Eastern University is a good deal higher than the total scores of the other groups. This may represent selection factors in terms of who goes on to graduate education. The graduate students are a higher achieving group in the arts and science areas as well. While this difference may be a function of an inadequate sample, it may also indicate that the graduate students are more knowledgeable at the one institution.

Also the eastern university students are slightly higher on all comparable scores than the California State College. The eastern school is usually viewed as a more competitive college in terms of types and numbers of students seeking graduation requirements. However the California State College is a newer school and has served as a senior college for most of its existence. This means that the junior colleges in the area have channelled their students into this school.

So far the scores have been dealt with according to specific countries. The most important conclusions will be derived from a cross-cultural comparison of scores in different societies. If all of the arts scores are placed together on one table and the science scores on another an overview of the relative achievements of the different schools is gained. Tables IV and V serve that purpose.

Discussion

At the top of the Arts Table we find the English University and the Public School. While at the lower end of the table are the Canadian first year University students. It seems reasonable to suggest that the mean scores reflect the level of cultural life within a particular college.

The mean scores may reflect the degree of urbanization of the institution in which the "Culture Test" was administered. The more rural locations offer less in the way of general cultural advantages. Further the whole approach to the "socialization" of the child differs in its lessened emphasis on the value of "arts" knowledge. Provincial values tend to stress practical information and a more vocational approach to the meaning of education. However,

we are less concerned with the explanation for the cultural level than with establishing the presence of the differences.

When the "science" scores are placed on a similar table we see that there are some changes in the position of the schools.

TABLE IV

MEAN "ARTS" SCORES OF GROUPS WITHIN THE DIFFERENT INSTITUTIONS

English University Department of Education	7.40
English Public School Sixth Form	7.30
American University Graduate Education	7.07
English University Department of Education	6.80
California State College—Graduate Education	6.10
English Grammar School Sixth Form (16-18 years old)	5.40
American University Undergraduate—Education	5.25
California State College—Juniors (Third Year)	5.10
English Women's Training College	4.70
English Women's Training College	4.40
Western Canadian B.Ed. Final and Fourth Year	3.87
English Women's Training College	3.20
Western Canadian 2nd Year Education	2.64
Western Canadian "1st Year Average"	1.97

TABLE V

MEAN "SCIENCE" SCORES OF GROUPS WITHIN THE DIFFERENT INSTITUTIONS

Western Canadian University B.Ed. Final Year	11.33
American University—Graduate Education	10.98
English Grammar School Sixth Form	10.90
English Public School Sixth Form	10.00
English University Education Department	9.30
California State College—Graduate Education	8.81
English University Education Department	8.60
American University—Undergraduate Education	8.12
Western Canadian 2nd Year Education	7.80
Western Canadian 1st Year Average	7.77
California State College—Junior Education	7.50
English Women's Training College	6.00
English Women's Training College	5.60
English Women's Training College	3.70

Discussion

There is no real difference between comparable Canadian and American students, while intending teachers from both countries have better scores than the women in the three English Training Colleges.

The very low scores at one of the English Training Colleges would indicate that these students are achieving at much lower levels than teacher trainees elsewhere. These students will probably be very limited in their ability to enhance the "science" knowledge of their pupils. One wonders about the criterion for selection in these institutions, and about the kinds of schools in which these teachers will work. Are these the elementary teachers who find positions in the poorer socio-economic school districts (8) ? Will their employment tend to limit the social and intellectual growth of their students?

An interesting difference between the "science" and "arts" scores is the much higher score gained by the Canadian Bachelor of Education student. He moves up from a very low "arts" score to the top position on the table of "science" scores. This represents three years of University study and some improvement over his first year "science" score. It may further indicate the value placed on scientific and useful vocational information which seems to be present in this institution. Students appear to view education as a functional process, useful for its value in preparing the individual for his vocational future.

Summary and Conclusion

A "Culture Test" developed by Kenneth Richmond in England was given to students in a Western Canadian University, a California State College and an Eastern American University. The aim of the "Culture Test" was to evaluate C. P. Snow's hypothesis about the conflict between the "Arts" and the "Science" culture.

This article was an attempt to compare educational institutions involved in the teacher training process, as well as comparing intending teachers to English Public and Grammar School students. The study did not attempt to measure C. P. Snow's ideas, but rather to utilize the "Culture Test" as a means of comparing levels of general knowledge in differing societies. Certain tentative conclusions may be drawn from these figures.

1. In recent years the importance of scientific training has been stressed in the schools. These tests results indicate that in Canada and America the students scored better on the science questions than on the arts questions. The differences may be caused by the popularity gained by science in the last decade. More scientific "facts" have entered into the teaching of science, and many new

courses and interesting approaches have been initiated. The traditional dominance of the "Arts" culture has been lessened and perhaps placed in a secondary position. It would appear that the "Arts" culture is in a very subordinate position in the Western Canadian area that was sampled in this study. This condition will probably continue unless the schools make a serious effort to improve this deficiency.

2. The schools differed along certain dimensions, namely whether they were rural or urban, and whether they were mass or élitist in their selection policy towards students. The English Public and Grammar Schools represented an élitist educational approach and their scores were among the highest. While the California College has more of a mass approach and their scores were poorer than the English schools.

A rural versus urban factor may also have been present in the very low "arts" scores obtained by the first and second year Canadian University students and the students in the Women's Training College in England.

3. There was a difference between teaching majors and non-teaching majors. Some of the lowest scores were obtained by the English women in the training colleges. In contrast the English Public School students, who do not generally become teachers, had higher scores. The Canadian "arts" scores were very low for individuals who will shortly serve as teachers in the public schools. The implication is simply that these individuals who are to serve as transmitters of the cultural heritage should be provided with broader awareness of the "arts culture".

Finally the authors wish to state that the sampling procedures, as well as the "Culture Test" utilized in this study, do not represent a carefully designed and controlled study. Further research is being conducted by Professor Richmond; and by the present authors toward that goal. This study should be viewed as a simple presentation of the mean scores obtained on Professor Richmond's "Culture Test" in somewhat comparable institutions. What these scores actually indicate is difficult to say unless a large number of variables are considered in the different societies (9).

REFERENCES

1. Snow, C. P., *Two Cultures and the Scientific Revolution,* Cambridge University Press, N.Y., 1959.
2. Riesman, David, *Constraint and Variety in American Education,* Doubleday Anchor Books, Garden City, N.Y., 1958, p. 79.
3. Richmond, Kenneth, "Two Culture Tests", Off Print from *New Saltire* No. 6, December 1962, p. 3.
4. Butts, R. Freeman, *A Cultural History of Education,* McGraw-Hill, N.Y. 1947, p. 571.

5. Key. V. O., Jr., *Public Opinion and American Democracy,* Alfred A. Knopf, N.Y., 1961, p. 343.

   "In the largest sense the educational system is a great mechanism for conserving the values, the institutions, the practices of the political order, as it inculcates from the primary school even through college the memories ,the unities and the norms of the system."

6. Richmond, Kenneth, "A Measure of Culture", *The Times Educational Supplement,* Friday, Septembr 29, 1961, p. 381.

   Examples of questions from the Arts and Science areas used by Mr. Richmond in his culture tests:

   1. *Arts.*
      The characters in "Under Milkwood" were—
      (a) English, (b) Breton, (c) Scots, (d) Welsh, (e) Irish, (f) French-Canadian, (g) none of these.
   2. *Science.*
      Which of these would you expect to be most interested in the mitochondrion—
      (a) a classisist, (b) an astronomer, (c) a philogist, (d) a physiologist, (e) an engineer, (f) none of these.

7. Halsey, A. H., Floud, Jean, Anderson, C. Arnold, *Education, Economy and Society,* The Free Press of Glencoe, 1961, p. 154-157.

8. *Ibid,* p. 539-543.

9. Richard, Kenneth, "Two Culture Tests", Off Print from *New Saltire* No. 6, December 1962.

   In this regard Mr. Richmond has said that: "In schools and colleges the hoohad over the two cultures has given rise to various makeshift proposals in which 'breadth', balance, and all 'roundness" figure as catchwords. All kinds of airy presuppositions are being made. Not least among them is the assumption that a broad based curriculum . . . is an essential prerequisite for breadth of mind in the learner . . . It was against this confused background that the idea of pioneering "Culture Tests" began to take root in my mind. . . . It seemed that an attempt had to be made to devise a yardstick of sorts, and to use it to obtain solid data in a field which was otherwise all too hazy. Inevitably, the tests so far devised have been crude affairs."

A FACTORIAL STUDY OF STUDENT ASSESSMENTS OF TEACHER PERFORMANCE*

B. E. J. McBride
Institute of Education
MacDonald College

The Problem

The evaluation of teaching proficiency has been the subject of intensive but unsystematic attack since Book's original study in 1905. As Anderson and Hunka (1963) have reported, the inter-rater reliability of teaching performance is low and, therefore, the value of say, Faculty of Education practice teaching ratings is dubious.

Theoretical Considerations

The Good Teacher: A Category

The problem of evaluating teaching performance is essentially one of identifying the good teacher. Categorization, as explained by Bruner (1951), is directly applicable to this problem, the concept, "good teacher" being a category, the identifying or criterial attributes of which must be based on value judgments. Since systems of values vary from individual to individual, the category "good teacher" must necessarily be a disjunctive one. (Bruner, 1956) In a disjunctive category there is no core of universal common factors; in fact, the identifying attributes are likely to bear little similarity to one another and category members to differ markedly. The consequences of the persistent but unsuccessful effort to change the category "good teacher" to a conjuctive one, wherein the criterial attributes are well known and undisputed, are serious. Researchers proceed as if there were a concept 'a good teacher' whose criterial atrtibutes are known to all and manifested in every teaching situation. This is not the case.

We might inquire what determines that which is characterized as good teaching? How does one arrive at a definition of good teaching? As Bruner relates, perception occurs on the basis of some expectancy or an hypothesis, nebulous though it may be. We not only see but we look for, not only hear but listen to. In short, perceiving takes place in a "tuned organism". (Bruner, 1951, pp. 123-124) The tuning is largely a product of past experience and training. Conceptual systems, described by Harvey, Hunt and Schroder (1961, p. 204) as the "immediate antecedents of cognitive

*The author is indebted to Dr. C. C. Anderson, under whose supervision the study was conducted.

and behavioral outcomes," develop, stimuli being perceived and interpreted as either refuting or confirming these conceptual systems.

The evaluator's conceptual system determines which aspects of a teaching performance he perceives as being the criterial attributes of "good teaching". He has, in effect, a set of established standards to which he believes teaching must conform. In an effort to preserve this conceptual system, he attends to aspects of input which reinforce it and screens out dissonant information. This is particularly true of people with closed minds as Rokeach (1960) suggests.

> The basic characteristic that defines the extent to which a person's system is open or closed, is namely, the extent to which the person can receive, evaluate, and act on relevant information received from the outside on its own intrinsic merits, unencumbered by irrelevant factors in the situation arising within the person or from the outside. (1960, p. 57)

It seems reasonable to assume that an adult will be more susceptible to such closemindedness than the child, who, in his lesser sensitivity to the socially desirable attributes of perception, should have a more flexible value system. The child's evaluation of teaching, therefore, will be less likely to be affected by defensive distortion, and accordingly it merits research attention.

Student Evaluations of Teaching Performance

The opinion that student evaluation of teacher performance deserve attention is by no means new. One need only consider that the student is a 'constant consumer'; that his judgment is not based on the somewhat false situation created when a supervisor enters the classroom, and that his evaluation is tempered by those of other students experiencing the same teaching, to recognize the likelihood of such ratings' being credible sources of information. The reliability coefficients attributed student ratings vary from .55 to .95. (Guthrie, 1927; Boardman, 1930; Root, 1931; Wilson, 1932; Bowman, 1934)

Student evaluation of teaching is not, however, without opposition. The student, it is pointed out, is not trained to observe teaching (perhaps this is the lack of bias valued by supporters of student rating). It is suspected that he may be influenced by the degree of his own success, that his sex may determine the favorability of his attitude towards teachers of one sex or the other, and that his age may affect his ability to rate. Also, it is asserted that the students' rating teaching is damaging to teacher morale and to student-teacher relations. Most of these charges have been investigated. Student success has been demonstrated to bear no relationship to student evaluation of teaching proficiency, (Remmer, 1930; Bowman, 1934) and, student rating, rather than damaging teacher morale, has been found to promote better teaching and

increased teacher self-understanding. (Bowman, 1934; Ward, Remmer and Schmalzried, 1941; Bryan, 1942; Amatora, 1951; Savage, 1959) Factors such as the age and sex of student raters, however, are considered more likely to affect ratings, although to a limited degree (Davenport, 1944). Another study (Morsh and Wilder, 1954), reports that the student factors, class size, sex, age, intelligence and mental age have little bearing on the evaluation made by students.

The Relationships Between Student Evaluation and Other Criteria of Teacher Proficiency

The findings regarding the efficiency with which students rate teaching are indeed diverse. An attempt to validate these ratings might be made through comparing them with the following criteria of teaching effectiveness, each of which is considered in turn: pupil growth scores, supervisory ratings, the accuracy of teacher perception of student body opinion, student rating of classroom interaction, and teacher self-ratings.

Pupil growth is perhaps the most widely recognized criterion of good teaching. Even so, limitations are many, for such growth may not be entirely the product of teacher effect though pupil intelligence and initial achievement be partialed out. Other factors remain uncontrolled. (Barr, 1953, p. 8) Some lie within the pupil in the form of general and specific abilities, previous training and motivation; others such as socio-economic status, respect for education and assistance by family members are aspects of the home environment; and still others like morale, physical facilities and the influence of other teachers might be attributed to the school and community. The difficulty of measuring pupil growth in its fullest sense, encompassing the development of personality, mental health, emotional maturity, and aesthetic interests, further complicates the problem. The few studies based on the pupil growth criterion present conflicting findings, some (Boardman, 1930; Lins, 1946; McCall, 1952; Guthrie, 1953; Keller and Clark, 1954) apparently laying a foundation for a relationship between student ratings of teaching and pupil growth scores, others (Jensen, 1961; Schmid, 1961) asserting that pupil gain bears no relationship to rating factors. Further study of pupil rating and the use of improved growth measures might lead to the solution of the problem.

The correlations between supervisory ratings and student ratings of teacher performance are frequently significant and positive as illustrated in the accompanying table. (Table I) A more recent investigation carried out in this province (Anderson, 1962) reveals contingency coefficients significant at the .05 and 0.1 levels respectively between pupil and supervisory and pupil and Faculty of Education raters. (Table II)

Table I

CORRELATION OF STUDENT RATING WITH ADMINISTRATIVE RATING OF INSTRUCTOR EFFECTIVENESS

Author	No. of. Teachers	Year	Level	Type of Relationship	Degree of Relatio ship
Knight	39	1922	elem. and high school	supervisory rating and student ranking	.74
Boardman	88	1928	high school	supervisory and student rankings	.56
Light	28	1930	high school	supervisory rating and student ranking	"close agreement"
Greene	32	1933	college	ratings by dean, peers, and students	All but 1 of dean's first 10 and 2 of peers' first 10 same as students, first 10
Bowman	30	1934	student	critic teachers' and student ratings	—.50 to .47
Starrak	entire faculty	1934	college	educational specialist and student graphic rating	75% of students showed close agreement; 25% divergence
Bryan	22	1937	senior high	administrative and student ratings	.36
	41		junior high	administrative and student ratings	.53
	22		senior high	administrative and student ratings	.36
	41		junior high	administrative and student ratings	.55
Brookover	33	1940	high school	supervisory and pupil ratings	—.08
	12		high school	av. 2 supervisory ratings and pupils ratings	.63
Ward, Remmers Schmalzried	40	1941	student	supervisory and student ratings	.875
Porter	27	1942	student	supervisory and student ratings	"close agreement"
Brookover	66 all male	1945	high school	supervisory and student ratings	"slight positive relationship"
Von	50	1946	high school	educational specialists' rating based on interview and autobiography and student ranking	none of 27 r's significant at .01 level. Range —.06 —.30, 28.
Lins	58 all female	1946	high school	composite supervisory rating and student ranking	

TABLE II

CONTINGENCY COEFFICENTS AMONG VARIOUS CRITERIAL INDICES OF TEACHING PROFICIENCY

Pupil	AB	Sup.	Pr. T.	Univ. Ave.
Pupil	.59**	.33*	.40**	.24
AB		.33*	.18	.17
Sup.			.21	.15
Pr. T.				.15

Pr. T.=Practice teaching marks given to the students while they were in training in the Faculty of Education in Edmonton, (1958-1959).

Univ. Ave.=Average mark in all other courses in the Faculty of Education in 1958-1959.

AB=In the field ratings by Anderson-Black-Buxton in the spring, 1960, after the students had almost completed one year of teaching.

Pupil=Questionnaire ratings by students of these teachers (grades 4-9) gathered by Anderson-Black-Buxton during their visit to the teacher and obtained while the teacher was outside the room.

Sup.=Superintendents' estimates gathered in the spring-summer of 1960.

*Significant at the .05 level of confidence.
**Significant at the .01 level of confidence.

[1] Reproduced from Anderson, C. C., (1962)

Similarly, the findings with regard to the relationships between teacher perception of student opinion or affect and student ratings of teaching tend to support student ratings. Indeed, the most frequently mentioned characteristic of the ideal teacher is possibly the possession of a keen understanding of students. The study which lends most direct support to this point of view is that by Gage and Suci (1961), in which a significant relationship was demonstrated between the degree to which high school teachers were able to estimate percentages of various student responses to a number of controversial questions and the 'positive affect' elicited by these same teachers in the student body. Further support is found in the work of Gage, Leavitt and Stone (1955), and to a certain degree in studies of leadership, the more successful leader (here, the teacher) being better able to predict the feelings and consequent action of group members and accordingly adjust his or her own behavior so as to produce harmony and direct positive affect towards himself. Steiner (1955) suggests that the leader, on the contrary, need be concerned with the hierarchy of the organization rather than individual idiosyncracies. Also, one must recognize that possible accurate leader perception of the feelings and attitudes of others and the favorable impressions of the perceptive leader on the part of his followers may both be the products of a third influence.

Research attention has recently been directed to the interaction between teachers and pupils. Drawthrone (1954), one of the few to explore the relationship between pupil ratings of teacher-pupils interaction and pupil ratings of teacher effectiveness, reports a correlation of .557 between the two; while Brookover (1940), as the result of an earlier and similar study reported a corresponding correlation of .639. A recent study by Gold (1962) related teacher effectiveness to the number of isolates in the class, teachers revealing a greater degree of dominative type thinking (measured on the F scale), being found to have a significantly greater number of isolates in their classes. Gold's work bears some similarity to that of a number of others, all of which lend it support. (Anderson, 1939, 1943; Anderson and Brewer, 1945, 1946; Anderson, Brewer and Reed, 1946; Thelen, 1957, 1959; Cogan, 1956, and Flanders, 1958, 1959, 1961a, and 1961b)

Self-ratings, subject to all the limitations of introspection, are not often recognized as being of much importance in the evaluation of teacher performance. A study by Webb and Nolan (1955), in which a correlation of .55 is reported between self-ratings and pupil ratings, suggests that they might be considered to lend some support to the use of student observations as a means of evaluating teaching proficiency.

It would appear, thus, that pupil evaluations of teaching performance possess sufficient merit to warrant further investigation, and on the basis of this tentative conclusion, the study described in subsequent sections was carried out.

The Development of Hypotheses

As Gombrich (1960) so aptly states, evaluative reactions . . . testify to the constant scrutiny with which we scan our environment with one vital question—are you friendly or hostile, a 'good thing' or a 'bad thing'? (1960, p. 232) In evaluating input, the individual must focus on certain cues, (Berlyne's (1960) stimulus selection). These cues are therefore of immediate concern in any investigation of teaching. To what cues does the student attend?

Studies indicate that students are keenly aware of sincere, understanding and responsive behavior on the part of teachers. (Cotsonas and Kaiser, 1962; Gage and Suci, 1951; Gage, Leavitt and Stone, 1955; Flanders, 1958, 1959, 1961b, and Gold, 1962) The student can then, in his responses to questions pertaining to classroom proceedings and his own related attitudes, be expected to reveal an awareness of, and concern with teacher-pupil rapport, particularly in terms of freedom to actively participate in classroom activity and discussion. Attitudes toward such freedom might be expected to be generally favorable, since pupil participation is one

of the marks of the classroom in which the teacher reveals concern with the individual student and understanding of his needs.

Just as he is aware of his freedom to participate, so also the child may be aware of and respond to the constraining, disciplinary and restrictive aspects of his environment. Both at home and at school, disciplinary measures serve as directives toward what is considered to be desirable behavior.

The control of behavior by means of reward and punishment or approval and disapproval can be explained in terms of arousal induction and reduction. As Berlyne (1960) relates, the individual is constantly seeking a steady state of optimum arousal, (pp. 200-209), significant departure from which represents 'unpleasure'. Both internal and external cues continually disturb this balance. In an effort to produce certain behavior, significant others induce arousal and set up corresponding incentives, the reaching of which by suitable behavior may reduce arousal, thus establishing working habits. Other theorists also explain the establishment of behavior patterns through the use of punishment and reward or anticipation of either. (Mowrer, 1950; Davis, 1948; McClelland, 1961)

The following hypothesis, with respect to the dimensions of student response to classroom events, ensue: that factor analysis of student responses will reveal

1. A factor describing the extent of student participation in classroom activity, and
2. a factor describing student awareness of and concern with the disciplinary and restrictive facets of the classroom,
3. a third hypothesis, or rather likely experimental possibility would be concerned with the possibility of developmental changes in the strength and incidence of these factors.

The Procedure

As part of a larger teacher evaluation program carried out in the spring of 1960 by members of the Faculty of Education of the University of Alberta, student responses to twenty questions regarding classroom activity, impressions of classroom interaction and interpersonal relationships, and attitudes toward work and attendance were collected in the classrooms of teachers who had been enrolled in the Junior Elementary and Intermediate classes in the Faculty of Education at the University of Alberta in the 1958-1959 winter session. All the teachers in whose classrooms student responses were recorded had thus had the same amount of teaching experience and had experienced the same training. The responses of 2023 students from grades three to nine inclusive were factor analyzed to determine whether the variations revealed in twenty-one responses could be accounted for in terms of fewer categories.

In each grade and in groupings from multiple-grade classrooms, the factors isolated were studied with reference to the items contributing most to them, and, on this basis they were matched with similar ones from other groups and grades. Tucker's coefficient of congruence (Harman, 1960, pp. 257-259) was employed to determine the degree of relationship between each factor in each grade or group and the most similar one in each other grade or group.

The Findings

Five factors were found to be constant in all or most of the groups. They are discussed in order of importance in succeeding paragraphs.

The first of these factors is a general satisfaction factor indicative of approval of the class and its activities and individual pleasure at participating therein. The most important items in terms of the size of their loadings and thus their contribution to this factor: Do you like to be in this class? Are you proud to be in this class? The measure of positive affect might be compared with the latter of two dimensions advanced by Foa (1961) as explaining human relationships, that of love-hostility and tending, of course, in the direction on the continuum labelled "love". It is not unlike what Lawrence (1961, p. 196) refers to as spontaneous informal group structure and Anderson (1939) labels socially integrative behavior.

The second factor mirrors the restrictive or disciplinary aspects of the classroom and the typical problems of noise, time wastage, and rule-keeping along with the prevalence of teacher scolding. The following items are examples of those contributing to it: Do the pupils often make so much noise that it's hard to work? Do some pupils break class rules a lot?

The remaining three factors are not so clearly defined as the former two; nor do they appear in all the grades and groups.

One of these, the third factor, might best be described as a participation factor, the students in it having revealed concern with the extent to which they work in groups, plan activities and formulate rules to be obeyed. These items serve to define it: Did the pupils make up the class rules they are supposed to obey? Do you help plan what the class is going to do?

Factor four reveals a tendency for the student to question the instruction offered coupled with a desire to absent himself from it, as revealed by the following items: Do you ever feel like staying away from school? Do you ever wonder why you have to do what you are doing in school?

The fifth factor relates to the individual's doing his best or doing satisfactory work and is well illustrated in items as follows: Is your

work usually good enough? Do you always do your best in this class?

On the basis of these findings, the first hypothesis, that factor analysis of student responses to classroom events would reveal a factor describing the extent of student participation in classroom activity, receives some support. The various factors chosen as representing participation in the groups compared fail, however, to relate to one another sufficiently to warrant one's concluding that students in various grades respond consistently. Further research with more refined and elaborate measures of participation might locate and define this quality more exactly.

The second hypothesis, that an analysis of student responses to classroom proceedings would reveal awareness of and concern with the disciplinary and restrictive facets of the classroom is also supported, the factor being relatively stable, students having responded consistently in terms of their cognizance of this aspect of the classroom.

The third hyothesis relating to the possibility of there being developmental changes in the strength and incidence of factors receives little support in the form of clear evidence, although the following two trends emerge:

a. a relationship between sex, dissatisfaction with classroom instruction and consciously doing one's best work at school. Whereas boys tend to be dissatisfied to a greater extent and desirous of avoiding the disliked instruction, girls appear to continue to do their best work and try to please teachers, even in the later years of junior high school.

b. a tendency for students to become, with increasing age, to a greater extent group-oriented, older students expressing pride in classes to which they belong, whereas younger ones place greater emphasis on individual relationships.

In this study ground is broken for a more intense study of student responses, the dominant dimensions of student response having been more effectively established than hitherto.

REFERENCES

Amatora, Sister M. Diagnostic teacher rating. *Catholic Ed. Rev.*, 1951, 49, 675-680.

Anderson, C. C. The evaluation of teaching proficiency: research findings and theoretical considerations. Unpublished manuscript, 1962.

Anderson, C. C. and Hunka, S. M. Teacher evaluation: some problems and a proposal. *Harv. Ed. Rev.*, 1963, 33, 74-96.

Anderson, H. H. Domination and socially integrative behavior. In Roger G. Baker, Jacob S. Kounin, and Herbert F. Wright (eds.) *Child behavior and development*, New York: McGraw-Mill, 1954, 459-483.

Anderson, H. H. Measurement of domination and of socially integrative behavior in teachers' contacts with children. *Child Dev.*, 1939, 10, 73-89.

Anderson, H. H. and Brewer, Helen M. Studies of teachers' classroom personalities, I. Dominance and socially integrated behavior of kindergarten teachers. *Applied Psychol. Monogr.*, 1945, 6.

Anderson, H. H. and Brewer, J. E. Studies of teachers' classroom personalities, II. Effects of teachers' dominative and integrative contacts on Children's classroom behavior. *Applied Psych.* Monogr., 1946, 11.

Anderson, H. H., Brewer, J. E., and Reed, M. F. Studies of teachers' classroom personalities, III. Follow-up studies of the effects of dominative and integrative contacts on children's behavior. *Applied Psych. Monogr.*, 1946, 11.

Barr, A. S., et al. Second report of the committee on criteria of teacher effectiveness. *J. Educ. Res.*, 1963, 46, 641-658.

Berlyne, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960.

Boardman, C. W. Analysis of pupil ratings of high school teachers. *Educ. Admin. Super.*, 1930, 16, 440-446.

Book. W. F. The high school teacher from the pupils' point of view. *Pedag. Seminar,* 1905, 12, 239-288.

Bowman, E. E. Pupil ratings of student teachers. *Educ. Admin. Super.*, 1934, 20, 141-147.

Brookover, W. B. Person to person interaction between teachers and pupils and teaching effectiveness. *J. Educ. Res.*, 1940, 34, 272-287.

Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey. (Eds.) *Perception: an approach to personality.* New York: Ronald, 1951, 121-147.

Bruner, J. S., Goodnow, J. J. and Austin, G. A. *A study of teaching.* New York: Wiley and Sons, 1956.

Bryan, R. G. Benefits reported by teachers who obtained written student reactions. *Educ. Admin. Super.*, 1942, 28, 69-75.

Cogan, M. L. Theory and design of a study of teacher-pupil interaction. *Harv. Educ. Rev.*, 1956, 26, 315-342.

Cotsonas, N. J. Jr., and Kaiser, H. F. A factor analysis of students' and administrators' ratings of clinical teachers in a medical school. 1962, *J. Ed. Psychol.*, 53, 219-223.

Davenport, K. *An investigation into pupil rating of certain teaching practices.* Further Studies in Attitudes, Series VIII, Studies in Higher Education XLIX, Lafayette, Indiana: 1944.

Davis, A. *Social-class influences upon learning.* Cambridge: Harvard University Press, 1948.

Drawthorne, C. L. Relationship between pupil and student-teacher interaction and pupil ratings of teacher effectiveness. *Educ. Admin. Super.* 1954, 40, 283-286.

Flanders, N. A. Analyzing teacher behavior. *Educ. Leadership,* 1961, b, 19, 173-180.

Flander, N. A. Interaction analysis: a technique for quantifying teacher influence. *Research Mem.* BER-61-2, April, 1961a.

Flanders, N. A. Personal-social anxiety as a factor in experimental learning situations. *J. Educ. Res.*, 1961, 45, 100-110.

Flanders, N. A. Teacher influence in the classroom: I. Research on Classroom Climate. Unpublished manuscript, 1958.

Flanders, N. A. Teacher-pupil contacts and mental hygiene. *J. Soc. Issues,* 1959, 15, 30-39.

Foa, U. G. Convergences in the analysis of the structure of interpersonal behavior. *Psychol. Rev.*, 1961, 68, 351-353.

Gage, N. L., Leavitt, G. S., and Stone, G. C. Teachers' understanding of their pupils and pupils' ratings of their teachers. *Psychol. Monogr.*, 1955, 69, No. 21, (whot No. 406).

Gage, N. L. and Suci, G. Social perception and teacher-pupil relationships. *J. Educ. Psychol.*, 1951, 42, 144-152.

Gold, H. A. The classroom isolate: an additional dimension for consideration in the evaluation of a quality education program. *J. Exp. Educ.*, 1962, 31, 77-80.

Gombrich, E. M. On physiognomic perception. *Daedalus. Winter,* 1960, 228-241.

Guthrie, E. R. The evaluation of teaching. *Training analysis and development of information.* Bulletin, 1963, 4, 199-206.

Harman, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1960.

Jensen, L. E. An non-additive approach to the measurement of teacher effectiveness. In A. S. Barr et al. *Wisconsin studies of the measurement and prediction of teacher effectiveness. J. Exp. Ed.*, 1961, 30, 70-87.

Keller, R. J. and Clark, K. E. *Student evaluation of classroom teaching. Bureau of Ed. Res.,* Univ. of Minnesota, March, 1954, (mimeographed).

Lawrence, P. R. *Organization, behavior, and administration.* Harewood, Illinois: The Dorsey Press and Richard D. Irwin, 1961.

Lins, J. E. The prediction of teaching efficiency. *J. Exp. Ed.* 1946, 15, 2-60.

McCall, W. A. *Measurement of teacher merit.* Raleigh, North Carolina: Superintendent of Public Instruction, No. 284, 1952.

McClelland, D. C. *The achieving society.* New York: D. Van Nostrand Company, Inc., 1961.

Morsh, J. E. and Wilder, Eleanor, W. *Identifying the effective instructor: a review of the quantitative studies,* 1900-1952. Chanute Air Force Base, Illinois: Training Aids Research Laboratory, Air Force Personnel, and Training Research Centre, Research Bulletin, October, 1954.

Mowrer, O. H. *Learning theory and personality dynamics.* New York: The Ronald Press Company, 1960.

Remmers, H. H. To what extent do grades influence student ratings of instructors? *J. Ed. Res.* 1930, 21, 314-317.

Rokeach, M. *The open and closed mind.* New York: Basic Books, 1960.

Root, A. R. Student ratings of teachers. *J. Higher Educ.,* 1931, 2, 311-315.

Savage, Marjorie L. Changes in student teachers through the use of pupil ratings. Unpublished Ph.D. Dissertation, University of Illinois, 1959.

Schmid, J. Factor analysis of the teaching complex. In A. S. Barr et al. Wisconsin studies of the measurement and prediction of teacher effectiveness. *J. Exp. Ed.* 1961, 30, 58-87.

Steiner, I. D. Interpersonal behavior as influenced by accuracy of social perception. *Psychol, Rev.*, 1955, 62, 268-274.

Thelen, H. A. Experimental research towards a theory of instruction. *J. Educ. Res.*, 1957, 45, 89-93.

Thelen, H. A. Work-emotionality theory of the group as an organism. In S. Koch (Ed.) Psychology: a study of a science. Vol. III, *Formulations of the person and the social context...* New York: McGraw-Hill, 1959, 544-611.

Ward, W. D., Remmers, H. H., and Schmalzried, N. T. The training of teaching personality by means of student ratings. *Sch. and Soc.,* 1941, 53, 189-192.

Webb, W. B. and Nolan, C. Y. Student, supervisor, and self-ratings of instructional proficiency. *J. Educ. Psychol.* 1955, 46, 4-47.

Wilson, W. R. Students rate teachers. *J. Higher Educ.,* 1932, 3, 75-82.

ORGANIZATIONAL PLANS FOR TEACHING PHYSICAL EDUCATION IN ELEMENTARY SCHOOLS

S. Boyd Taylor and Walter H. Worth

Introduction

Traditionally, organization for staff utilization in the elementary school has centered around the idea of the self-contained classroom wherein one teacher instructs the pupils in all of the studies of a particular grade. Although widely used, it has been recognized that the self-contained classroom has certain weaknesses. Its use requires that each teacher must be capable of teaching all subjects, it tends to ignore the factors of aptitude and interest on the part of the teacher, and it makes no attempt to capitalize on special talents that may exist within a staff. Because of these weaknesses a number of alternatives to the self-contained classroom type of organizational plan have been proposed. These alternatives usually consist of some form of departmentalization, or the use of specialist teachers in subjects like art, music, and physical education.

In recent years, the centralization of rural school facilities and the rapid growth of population in most urban centers has meant that both rural and urban school systems are becoming larger and more complex. Increased size has, generally, meant larger school staffs, better facilities, and in some cases the availabilty of specialist services. In order to provide for the best possible type of instruction in large elementary schools, and in order to make effective use of facilities and available specialist and supervisory help, a re-examination of organization for instruction for the elementary school program as a whole, and for individual subjects, needs to be made.

Physical education is one subject which, due to the lack of qualifications or desire on the part of many teachers, may not always receive the attention it should under the self-contained classroom type of organization. In regard to organization for instruction in physical education, an elementary school may choose one of four general policies: (1) have each classroom teacher teach physical education to his own class, (2) make allowances for teachers to exchange classes so that those with special training and/or interest in physical education may relieve those who have not, (3) have physical education taught by a specialist who is a member of the school staff, and (3) have physical education taught by a visiting specialist employed by the unit or school system.

Purpose

The study sought to accomplish the following five purposes: (1) to identify existing organizational provisions for teaching physical education in the elementary schools of Saskatchewan, (2) to determine the type of organizational plans preferred by elementary school teachers and administrators; (3) to determine the extent to which elementary school teachers feel competent in teaching physical education; (4) to compare the organizational plans used in large and small schools and in city and non-city school; and (5) to determine whether present practice, preferred practice, and teachers' feelings of competence are significantly related to the teacher variables of age, sex, and professional preparation in physical education.

Procedure

Collection of Data

In order to gather information from principals and teachers for the study, two types of questionnaires, one for principals and one for teachers, were constructed and distributed to a randomly selected sample of thirty city schools and thirty non-city schools in the province of Saskatchewan. Only those schools having eight or more elementary classrooms in operation were included in the population studied. Responses were received from 57 principals and 588 teachers. This represented a 95 per cent return from principals and an 89 per cent return from teachers.

The teacher questionnaire asked for personal data about the respondents, such as age and sex, the grade level they were teaching, and the amount of professional preparation for teaching physical education. In order to determine the type of organizational plan presently being used for instruction in physical education, and the type of plan teachers would prefer for instruction in physical education, teachers were requested to check one item in each of the following sections of the questionnaire:

I. *Present Practice in Teaching Physical Education*
   (1) Teach physical education to my own class.
   (2) Physical education is being taught to my class by another elementary staff member.
   (3) Physical education is being taught by a special physical education instructor who is a member of the elementary school staff.
   (4) Physical education is being taught by a visiting specialist employed by the unit or school system.
   (5) No physical education is taught in my class.
   (6) Other arrangements. (Specify).

II. *Preference in Regard to Organization for Instruction in Physical Education*
   (1) Physical education taught by the classroom teacher.
   (2) Physical education taught by a classroom teacher in the school as part of a system of departmentalization or semi-departmentalization in which teachers "trade" subjects.
   (3) Physical education taught by a special physical education instructor who is a member of the school staff.

(4) Physical education taught by a visiting specialist employed by the unit or school system.

III. *Competence in Teaching Physical Education*

(1) Feel proficient in teaching all areas of the physical education program.
(2) Feel proficient in teaching most areas of the physical education program.
(3) Feel inadequate in teaching most areas of the physical education program.
(4) Feel inadequate in teaching all areas of the physical education program.

The principal questionnaire elicited basic data about the school. It asked principals to indicate whether any form of departmentalization or semi-departmentalization was being used for physical education instruction at different grade levels in the school. Principals were also requested to indicate the type of plan they would prefer for instruction in physical education at different grade levels in their school. The possible plans were the same as those listed in section II of the teacher questionnaire.

Treatment of Data

The sample studied was divided into sub-groups according to selected characteristics. The characteristics investigated were size of school, grade level that the respondents were teaching, and the teacher's age, sex, and amount of professional preparation for teaching physical education. Table I shows the size of the resulting sub-groups.

TABLE I

NUMBER OF RESPONDENTS CLASSIFIED ACCORDING TO SIZE OF SCHOOL, GRADE LEVEL TAUGHT, AGE, SEX AND AMOUNT OF PROFESSIONAL PREPARATION FOR TEACHING PHYSICAL EDUCATION

Characteristics	Principals N	Teachers N
Size of School		
Small (eight to twelve rooms)	40	334
Large (Thirteen or more rooms)	17	254
Grades Taught		
1-3	.....	237
4-6	.....	218
7 and 8	.....	133
Age		
Under thirty	4	286
Thirty to forty-five	31	175
Over forty-five	22	127
Sex		
Female	5	455
Male	52	133
Professional Preparation in Physical Education		
No classes	16	181
One class	24	346
Two classes	12	47
Three or more classes	5	14

The significance of differences among the sub-groups in regard to present practice for teaching physical education, preferred plans, and expressed feelings of competence in teaching physical education were calculated by chi-square tests of independence. When the value of chi-square was found to be large enough to indicate a level of probability of .05 or lower ($p \geqq .05$), the difference was considered significant for the purposes of this study.

Findings

Present Practice in Teaching Physical Education

1. Approximately 79 per cent of the teachers in the sample schools were teaching physical education to their own classes. In almost all cases where teachers were not teaching physical education to their own classes it was being taught by another staff member as part of a system of departmentalization or semi-departmentalization. Less than 1 per cent of the classes were taught by a special physical education instructor who is a member of the school staff and in no instance was physical education being taught by a visiting specialist. In approximately 2 per cent of the classes there was no physical education being taught.

2. Principal's responses indicated that in 16 per cent of the sample schools there was some departmentalization being used for physical education instruction in grades one to three, in 53 per cent of the schools this form of organization was being used to some extent in grades four to six, and in 63 per cent of the schools it was being used to some extent in grades seven and eight. Although these percentages indicate that there is some departmentalization being used in many schools, teachers' responses showed that only 5 per cent of the teachers in grades one to three, 25 per cent in grades four to six, and 32 per cent in grades seven and eight were having their physical education taught by another staff member. These differences between grade levels were significant, indicating that there is more departmentalization or semi-departmentalization being used in the upper elementary grades.

3. The type of organizational plan being used was significantly related to both size and location of school. In larger schools there was a higher percentage of teachers who were using a plan whereby physical education was being taught by another staff member and there was a higher percentage of teachers using this type of organization in non-city schools than in city schools.

4. There was a significant relationship between the organizational plans used by teachers and the age of the teachers. There was a tendency for older teachers, in the schools studied, to have their physical education taught by another staff member.

5. There was no significant relationship between the type of organizational plans being used and the sex of the teachers.

6. Approximately 31 per cent of the responding teachers had taken no courses in physical education as part of their professional preparation, about 59 per cent had taken one course, 8 per cent had taken two courses, and slightly more than 2 per cent had taken three or more courses. There was a significant relation between the type of organizational plan being used by teachers and the number of physical education courses that teachers had taken. Teachers with more professional preparation for teaching physical education were more likely to be teaching physical education to their own classes, and possibly to classes other than their own.

7. Approximately 11 per cent of the teachers in the sample schools were teaching physical education to two or more classes. There was a significant relationship between the number of classes teachers were instructing in physical education and the teachers' age, sex, and amount of professional preparation in physical education. Of the teachers who were instructing two or more classes in physical education the majority were male, were under thirty years of age, and had taken two or more professional physical education courses.

Organizational Plans Favored by Principals and Teachers

1. The majority of principals indicated that they would prefer to have physical education taught by the classroom teacher in grades one to three, but would prefer a system of departmentalization or semi-departmentalization in grades four to six, and would like to see physical education taught to grades seven and eight by a special physical education instructor in the school.

2. The majority of teachers indicated that they would prefer to have physical education taught to grades one to three either by the classroom teacher or by a special physical education instructor in the school. For grades four to eight the majority of teachers preferred to have physical education taught by a special physical education instructor in the school.

3. There was a significant relationship between the sex of the teacher and the organizational plans favored for instruction in physical education. Although the majority of both male and female teachers favored an organizational plan whereby physical education could be taught by someone in the school other than the classroom teacher, a higher percentage of male teachers than female teachers preferred this type of organization, with a fairly large number of female teachers preferring to have physical education taught by the classroom teacher.

4. There was no significant relationship between the plans favored by teachers and the age of teachers.

5. There was no significant relationship between the plans favored by teachers and the amount of professional preparation that teachers had taken in physical education.

6. There was a significant relationship between the type of organizational plans preferred by teachers and the type of plan teachers were actually using. Although the majority of teachers preferred a plan whereby physical education is taught by a special physical education instructor in the school, teachers not choosing this type of organization tended to choose the type of plan that they were actually using at the time of the survey.

Feelings of competence in Teaching Physical Education

1. Approximately 8 per cent of the teachers in the sample schools indicated that they felt competent in teaching all areas of the physical education program; 53 per cent felt competent in teaching most areas of the physical education program; 34 per cent felt inadequate in teaching most areas of the program; and 5 per cent felt inadequate in teaching all aspects of the physical education program.

2. There was a significant relationship between the degree of competence expressed by teachers and the age of the teacher. Approximately 66 per cent of those under thirty years of age indicated that they felt competent in teaching most or all of the physical education program, while 57 per cent of the teachers in the thirty to forty-five age group, and 55 per cent of those who were over forty-five years of age indicated that they felt competent in teaching most or all of the physical education program.

3. There was no significant difference between male and female teachers in the degree of competence expressed towards the teaching of physical education.

4. There was a significant relationship between the degree of competence expressed by the responding teachers and the number of professional physical education courses that they had taken. The more professional preparation teachers had taken in physical education, the more competent they tended to feel in teaching physical education.

5. There was a significant relationship between the degree of competence expressed by the responding teachers and the type of plan that these teachers were using for instruction in physical education. Teachers who felt competent in teaching most or all of the physical education program were more likely to be teaching physical education to their own classes, while those who felt inadequate in teaching physical education were more likely to have

physical education taught to their classes by another staff member, or to have no physical education taught to their class.

6. There was a significant relationship between the degree of competence expressed by teachers and the type of plan that teachers preferred for instruction in physical education. Most of the teachers who preferred to have physical education taught by the classroom teacher were teachers who felt competent in teaching most or all of the physical education program. while only about one-half of those who preferred a type of organization in which physical education could be taught by another staff member, were teachers who felt competent in teaching most or all of the physial education program.

Conclusions and Implications

An examination of the above findings gives rise to a number of conclusions and implications relative to physical education instruction in elementary schools.

Organizational Provisions for Instruction in Physical Education

This study indicates that the majority of principals and teachers in Saskatchewan's elementary schools would prefer a type of organizational arrangement for instruction in physical education other than the one which they are presently using. While it does not necessarily follow that existing arrangements are totally unsatisfactory, this does mean that many teachers are dissatisfied with present conditions and that both principals and teachers believe that physical education might best be taught using other organizational arrangements. Lacking objective evidence to indicate which type of organization is most suitable at different grade levels, the opinion of the majority of principals and teachers is probably sufficient reason to suggest that some changes should be made.

It was shown in the study, that approximately one-third of the teachers in elementary schools have had no professional preparation for teaching physical education and that over one-third of the teachers do not feel sufficiently competent in teaching physical education. It seems unrealistic to continue to expect these teachers to teach physical education adequately.

The study also indicates that in those schools where arrangements have been made for teachers to have their physical education taught by another staff member, the persons assuming responsibility for physical education tend to be better qualified and more competent than the teachers they are relieving. Having physical education taught by teachers who are better qualified and who feel competent in teaching this subject should result in students receiving better physical education instruction.

Both the opinion of principals and teachers as expressed in this study, and the literature on physical education, suggest that in the primary grades physical education should be taught by the classroom teacher. This study reveals, however, that most principals prefer a system of departmentalization or semi-departmentalization for instruction in physical education in grades four to six, while most teachers would prefer to have physical education taught by a special physical education instructor at this level, and that both principals and teachers prefer to have physical education taught to grades seven and eight by a special physical education instructor in the school. The relatively small size of many elementary schools and the lack of qualified physical education personnel means that, in most cases, it is unrealistic to hope to have physical education taught by a special instructor. It is possible, however to make provisions for teachers to trade subjects so that those with ability and interest in physical education may relieve other teachers of their responsibilities in this subject. It would seem most advisable then to have physical education taught by the classroom teacher in the primary grades and to make provision for, and encourage, teachers to trade subjects in grades four to eight so that physical education may be taught by those who are most capable of giving instruction in this subject. In large elementary schools, where it is feasable to do so, a teacher fully qualified to teach physical education could be employed to teach classes in grades four to eight and to give assistance in this subject to the teachers of grades one to three. A system of team teaching with one member of the team being highly qualified in physical education might also be used to advantage in the upper elementary grades.

In view of the fact that at the present time most teachers are expected to teach physical education to their own classes, and since this situation is likely to continue to exist for some time in many schools, every effort should be made to provide in-service help in physical education. This means that qualified personnel should be employed to act as supervisors or consultants in units and school systems. With the organization of larger administrative units and the consolidation of attendance areas this should be possible in most cases.

Teacher placement

This study indicates that there are a considerable number of teachers in the schools who have taken two or more professional physical education courses. There are, in fact, enough of these teachers in Saskatchewan so that, theoretically, one could be placed in every school having eight or more elementary classrooms in operation. In addition to those teachers who have taken extra pro-

fessional preparation in physical education there are probably many teachers who because of background or interest, are very capable in teaching physical education. Through careful selection and placement of teachers on elementary school staffs it should be possible to have in each school at least one person who is capable of relieving teachers who cannot teach physical education satisfactorily.

It is shown in the study that younger teachers tend to feel more competent in teaching physical education than older teachers. This also has implications for teacher placement. By balancing school staffs in regard to the age of teachers, administrators will increase the possibility of having at least some teachers who are willing and capable in the field of physical education.

Teacher Education

At the present time the majority of teachers in elementary classrooms are expected to teach physical education. As indicated by this study, many of these teachers have had no professional preparation for teaching this subject. As long as the present organizational arrangements in elementary schools exist, it will be essential that prospective elementary school teachers receive some preparation for teaching physical education. An increasing number of prospective elementary school teachers should also be encouraged to specialize in physical education so that they may assume physical education responsibilities in the upper grades of elementary schools. At the present time there are few people specializing in physical education and almost all of those who do become qualified in this subject are absorbed into high school staffs.

Further Research

There is a need for research that would evaluate and compare the physical education programs offered under different types of instructional organization. The development of criteria to evaluate physical education programs offered in different schools would make it possible to relate effectivess to organizational arrangements, as well as to additional factors such as facilities, teacher variables, supervisory assistance, and size and location of schools.

Similar studies should be carried out in other subject areas. The composite results of such studies might lead to the development of an organizational plan which would improve the effectiveness of the total educational enterprise at the elementary level.

PREDICTIVE VALIDITY OF THE METROPOLITAN READINESS TESTS

HARVEY W. ZINGLE
Faculty of Education

A. E. HOHOL
Superintendent of Jasper Place Public Schools

Introduction

In recent years the concept of readiness has come to occupy a prominent place in educational theory.

It is generally believed today that the assessing of a child's readiness for any aspect of learning is the first step that should be taken in successful teaching. The slogan "Begin where the child is" has gained great currency in educational practice and has, in many respects, revolutionized teaching methods and organizational plans in schools.

As the term readiness is used today it implies more than just emotional readiness, although that is important. It also includes physiological maturation, general mental ability, and experimental background or educational readiness. To be completely ready for an educational activity or learning experience, a child must want to learn, be sufficiently mature physiologically, possess appropriate mental abilities, and finally have had the right kind of educational experiences.

Thus, although most people agree that a child can profit sufficiently from attending school only when he has achieved minimum readiness, not all people agree (especially parents and school authorities) on an objective way of estimating readiness. Because of this inevitable conflict between parents and school authorities, most schools have in the past used the highly objective criterion of chronological age. The improvement of intelligence testing during the last couple of decades led most school systems to relax chronological age limits to permit the use of the mental age concept. Further sophistication in the field of tests and measurement has led many school systems to adopt a policy under which chronological age is combined with tests that attempt to measure the extent of the maturation of those skills considered most necessary for achievement in Grade I.

One such battery of tests is the Metropolitan Readiness Tests, which according to its authors (10) was devised to measure the traits and achievements of school beginners that contribute to their readiness for first-grade instruction.

Problem

The publisher (10) of the Metropolitan Readiness Tests, after stating that the validity of their tests may be judged in terms of the correlation of "Readiness scores" with first-grade achievement, cite the results of a study done in Lebanon County, Pennsylvania, involving 487 cases. All of the 44 pupils rated as "Superior" on the Metropolitan Readiness Tests administered in September were found to be above the national norm in average "reading achievement," and all but one were above this norm in "Numbers" when tested in the following February on the appropriate tests of the Primary I Battery of the Metropolitan Achievement Tests. Of 46 pupils rated "Poor Risk" on the Metropolitan Readiness Tests, 22 were found to be below the national norm in average "reading achievement" and 39 below in "Numbers" when tested in February.

Gardner (Buros, 1963) who quarrels with the conclusions drawn from this study, notes that even if the norm is accepted as a minimal level of competence, which is certainly a debatable issue, about half of those who were considered "unready" for reading actually reached the norm of achievement. An unbiased interpretation of these data certainly would emphasize far more than does the publisher's manual, the extremely tentative meaning of the Metropolitan Readiness Test scores for individual diagnosis. This is especially crucial in the lower range where a difference in raw score of one may mean non-admission to school as opposed to admission.

At the same time educators have found that using chronological age or mental age is certainly no better. According to Dey (6) neither chronological age nor mental age are completely reliable as measures of the optimal age at which all children can profit from formal instruction. Bevington (1) in a study of 640 pupils in the Edmonton Public School System, including a large proportion of pupils under the age of six, found no correlation between chronological age and achievement.

It is dubious, then, whether the presently used methods of establishing whether or not a child is ready for formal instruction are adequate. At the same time the administrative decision regarding the child's admittance to school is usually irrevocable. Recognizing that such decisions have to be made, and further that in many schools these decisions are made on the basis of subjective judgments has led the authors of this study to undertake the testing of the following hypothesis:

1. The Metropolitan Readiness Tests are good predictors of success in Grade I reading and arithmetic, as measured by year-end standardized tests and year-end teacher ratings.

Two subsidiary hypotheses of this study were:

1. Teachers' year-end ratings tend to discriminate against boys more than standardized tests do.
2. Teachers' year-end ratings tend to discriminate against older students more than standardized tests do.

Experimental Design

Sample

The subjects used in this study were all 545 of the Grade I students who were admitted to the Jasper Place Public Schools in the 1961-62 term. The first-grade readiness testing program of the Jasper Place Schools consists of the administration of all the Metropolitan Readiness Tests to all the candidates who seek first-grade admission. The legal school entry age in Jasper Place is six years, but any one who is at least five years and six months may write the tests. A total tests score of more than 40 on the Metropolitan Readiness Tests qualifies an under-age candidate for admission to Grade I. All children of legal age are admitted regardless of their test scores. Thus the age distribution of the Grade I students used in this study is negatively skewed. The actual distribution is shown in Table I.

TABLE I

DISTRIBUTION OF SUBJECTS BY CHRONOLOGICAL AGE

Age	5-6	5-7	5-8	5-9	5-10	5-11	6-0	6-1	6-2	6-3	6-4	6-5	6-6	Over 6-6	Total
	19	23	38	32	48	46	49	52	48	41	45	31	27	46	545

Procedure

Even though the Metropolitan Readiness Tests are used as a screening device for only the underage candidates it was administered to all candidates in June 1961. In June 1962 the following tests were administered to all Grade I students:

1. The Marion Monroe Reading Test.
2. The Edmonton Public School Board Grade I Achievement in Arithmetic Test.

In addition to obtaining these test scores the teachers were asked to rate each of the students on a five point rating scale. The ratings were defined as indicated in Table II.

Findings and Discussion

Intercorrelations of the predictor scores (Metropolitan Readiness scores) with the criterion scores (year-end achievement scores) for

the 545 subjects of this study are reported in Table III. Although the correlations are not as high as the correlations given in the test manual all are high and positive. To be fair to the test publisher, however, it must be pointed out that the children who made a total score below 40 were not admitted to school and so were not a part of the study. It is expected, therefore, that if these children had been retained the intercorrelations would have been somewhat higher.

TABLE II

TEACHER RATING SCALE

Rating	Percentile Equivalent
1	90-99
2	75-89
3	25-74
4	10-24
5	1-9

TABLE III

INTERCORRELATIONS OF METROPOLITAN READINESS TEST SCORES WITH YEAR-END MEASURES OF ACHIEVEMENT

Metropolitan Readiness Scores	Year-End Achievement Scores		
	Monroe Reading	Edmonton Arithmetic	Teacher Rating
Reading	.31	.45	.43
Number	.39	.53	.50
Total Score	.40	.56	.53

One rather interesting observation that can be made about the correlations in Table III is that the Metropolitan Reading Test predicts achievement in arithmetic (as measured by the Edmonton Arithmetic Test) significantly better than it does achievement in

reading (as measured by the Monroe Reading Test). This difference in correlations is significant at the .01 level of confidence.

Returning to the first hypothesis of this study, one would have to conclude that the Metropolitan Readiness Tests produce information which is useful when one attempts to make predictions regarding the academic success of Grade I students. In other words, the predictive validity of the Metropolitan Readiness Tests has been verified.

The first subsidiary hypothesis was stated as follows: Teachers' year-end ratings tend to discriminate against boys more than standardized tests do. This hypothesis was partially inspired by the fact that most studies (Samuels, 1943; Carrol, 1941; Donnelly, 1940) which show a sex difference in achievement at the Grade I level are based on teacher ratings. And one might suspect that the fact that first-grade teaching staffs are predominately female might be an ancillary factor of bias. It would appear, however, that since the point biserial correlation between sex and the Monroe Reading Test (—.11) and the point biserial correlation between sex and teacher rating (—.14) are both negative and significant this hypothesis is not upheld (Table IV). It would appear that there is a real sex difference in reading favoring the girls and also that this difference, which has been observed in the past, is not only a phenomenon of subjective teacher ratings but is also a phenomenon of standardized tests. It is also evident from an inspection of Table IV that this tendency to discriminate against boys is not a characteristic of the Edmonton Arithmetic Test ($r = .00$).

TABLE IV

POINT BISERIAL CORRELATIONS BETWEEN SEX AND YEAR-END MEASURES OF ACHIEVEMENT

	Year-End Achievement Scores		
	Monroe Reading	Edmonton Arithmetic	Teacher Rating
Sex	—.11	.00	—.14

The second subsidiary hypothesis was stated as follows: Teacher's year-end ratings tend to discriminate against older students more than standarized tests do. None of the correlations between age and year-end measures of achievement presented in Table V are significant. However, the correlation between age and Monroe Reading (0.6) differs significantly (.05 level of confidence) from the correlation between age and teacher rating (—.07). As the difference between the correlation of age with the Edmonton Arith-

metic Test is not significantly different from the correlation of age with teacher rating the second hypothesis is only partially supported.

TABLE V

INTERCORRELATIONS OF CHRONOLOGICAL AGE WITH YEAR-END MEASURES OF ACHIEVEMENT

	Year-End Achievement Scores		
	Monroe Reading	Edmonton Arithmetic	Teacher Rating
Age	.06	.01	—.07

Conclusions

1. The Metropolitan Readiness Tests are good predictors of success in Grade I reading and arithmetic as measured by year-end Standardized tests and year-end ratings. School administrators can no doubt find the Metropolitan Readiness Tests useful in helping them make admission decisions about pre-school children.
2. As both the Monroe Reading Test and the teachers' ratings tend to discriminate about equally against the boys it would appear that there is a real sex difference in reading.
3. The hypothesis that teachers' ratings tend to discriminate against older students more than standardized tests do was only partially upheld. The findings are very inconclusive as there seemed to be some indication of discrimination as far as reading is concerned but none in arithmetic.

REFERENCES

1. Bevington, W., "Effect of Age of Entrance into Grade One on Subsequent Achievement." Unpublished Master's Thesis, University of Alberta, 1957.
2. Brodley, B. E., "An Experimental Study of the Readiness Approach to Reading." *Elementary School Journal*, 46, 1956, pp. 262-267.
3. Buros, O. K. (Ed.) *The Fourth Mental Measurements Yearbook*, Highland Park, New Jersey, The Gryphon Press. 1953.
4. Carrol, M. W., "Sex Differences in Reading Readiness." Unpublished Master's Thesis, Boston University, 1941.
5. Dean, C. D., "Predicting First Grade Achievement." *Elementary School Journal*, 39, 1939, pp. 609-616.
6. Dey, J. D., *Theory and Practice Governing the Time of School Entrance.* Monographs in Education No. 4, The Alberta Advisory Committee on Educational Research, University of Alberta, Edmonton, 1960.
7. Donnelly, H. E., "A Study of Word Recognition Skills in Grade One." Unpublished Master's Thesis, Boston University, 1940.
8. Gates, A. I., "The Necessary Mental Age for Beginning Reading." *Elementary School Journal*, 37, 1937, pp. 497-508.
9. Hildreth, G., *Readiness for School Beginners.* Yonkers-on-Hudson, New York, World Book Co., 1950.
10. *Metropolitan Readiness Tests,* Yonkers-on-Hudson, New York, World Book Company, 1949.
11. Samuels, F. L., "Sex Differences in Reading Achievement." *Journal of Educational Research*, 36, 1943, pp. 594-603.

RELATION OF TEXTBOOK DIFFICULITY TO MATHEMATICS ACHIEVEMENT IN JUNIOR HIGH SCHOOL

L. Doyal Nelson

The importance of the role of the textbook in determining the content, organization, and mode of presentation of school mathematics courses can hardly be over-estimated. The production of textbooks in recent years with content strikingly different from that of traditional textbooks has been largely responsible for the current revolution in school mathematics. This change in content has brought with it the question of how the material can best be organized and presented to the students. In other words, is there any one pattern of organization or method of presentation of mathematics ideas through textbooks which will foster greater learning efficiency on the part of the pupil?

With our present knowledge of the nature of the learning process it is doubtful if we will find a definite answer to this question which would apply to all pupils. However, if there were two sets of mathematics textbooks which contained essentially the same material and which differed only in organization and method of presentation, it would be possible to shed some light on the question for pupils of certain levels of ability. This could be done by taking two groups of pupils of comparable ability, assigning one set of textbooks to one group of pupils, and the other set to the other group, and then comparing the mathematics achievement of these two groups of pupils after they had used the texts for a specified period of time.

One of the most influential organizations in the current revolution in school mathematics is the School Mathematics Study Group (SMSG). They have focussed their attention on the production of textbooks with content in line with the needs of pupils in a modern technological society. They have produced for both the grade seven and grade nine levels two sets of textbooks in which the content is similar but which differ in reading difficulty, organization, and method of presentation. The text *"Mathematics for Junior High School, Volume 1"* was prepared especially for college-bound seventh grade students. A modification in the presentation and organization of much of the material in this text was subsequently made for slower learning seventh grade pupils and is contained in the text called *"Introduction to Secondary Mathematics, Volume 1"*. According to an SMSG newsletter "The changes and adjustments which

were made were prompted by a desire to simplify the presentation and reduce the reading difficulty. Explanatory sections were shortened and exercises added to lead pupils through simple steps to appropriate conclusions[1]". These texts contain a modern treatment of such topics as systems of notation, the system of whole numbers, rational numbers, factors and primes, non-metric geometry, and so forth.

In a similar manner the textbook *"Mathematics for High School, First Course in Algebra"* was prepared for college-bound ninth grade pupils and the organization and presentation of its content was subsequently modified for slower learning pupils. The modified version appears in the textbook called *"Introduction to Algebra"*. These texts contain a modern treatment of such topics as: the system of real numbers, properties of operations and order, sets and sentences, polynomial and rational expressions, functions and the like. For a complete picture of the nature of the content and nature of the modifications the reader should refer to the textbooks themselves.[2]

Problem

It was the object of this investigation to study the following questions: What would be the effect on the mathematics achievement of high ability students who used the SMSG texts designed for slower learning pupils? Would their mathematics achievement be different from that of similar high ability students who used the SMSG sets designed for college-bound capable students?

Specifically, the basic hypotheses to be tested were as follows:

(a) There is no difference in the mathematics achievement, as measured by a standardized test, between high ability grade seven students who use the seventh grade SMSG text for college-capable students and those who used the text for slower learners.

(b) There is no difference in the mathematics achievement, as measured by various unit tests prepared specifically to test achievement in material covered by the texts, between high ability grade seven students who use the seventh grade SMSG text for college-capable students and those who use the text for slower learners.

Similar hypotheses were tested for high ability ninth grade pupils.

Design

This investigation was conducted in fourteen schools in Minnesota during the 1961-62 school term. Each of the schools provided two mathematics classes—both of which consisted of either high ability seventh grade pupils or high ability ninth grade pupils.

1 Newsletter No. 11, School Mathematics Study Group, Leland Stanford University, 1962; page 15.

2 These textbooks may be obtained from: Yale University Press, School Mathematics Study Group, 92A Yale Station, New Haven, Connecticutt.

One of the classes in each school used the SMSG text for college-capable students at the appropriate grade level and the other class in the school used the SMSG text which had been modified for slower learners.

Seventh grade pupils for the two experimental classes in each school were selected from among those grade seven students in the school who were above the mean in measured mental ability. Ninth grade pupils for the experiment were selected from among the top third in mental ability. Teachers and administrators in each school were asked to use official files and their knowledge of pupil ability to make up the two classes so that their mean abilities were as nearly alike as possible. Six pairs of seventh grade classes and eight pairs of ninth grade classes were obtained for the experiment.

One mathematics teacher from each school was assigned to teach both experimental mathematics classes. Teachers were instructed to use a method of presentation for each class which would conform closely to the textbook assigned. Every attempt was made to keep other variables as nearly constant as possible. For example, teachers for the experiment were chosen by the Minnesota National Laboratory and throughout the experimental period maintained close contact with members of this organization. Officials of the Laboratory met with the teachers of the experimental classes once each month, helped them with instructional problems, and consistently stressed the necessity of controlling such variables as time spent in developing ideas, time spent in review, standards in scoring tests, use of motivating devices, and so forth. The teachers involved were not typical of junior high school mathematics teachers. All of them had had previous experience teaching experimental programs, and all but one of them had taken formal classes specifically designed to help them to interpret modern programs in school mathematics. The total number of quarter credits they held in college mathematics varied from 24 to 67.

There was nothing in the design of this experiment to guarantee that a student who was classified as high ability in one school would be so classified in another. This lack of cross classification made it necessary to consider each pair of classes in each school to be a separate experimental situation for the purpose of testing the hypotheses. In the six pairs of seventh grade classes there were 285 students and in the eight pairs of ninth grade classes there were 460 students.

Various tests were administered during the course of the experiment and the resulting measures provided the basis for testing the hypotheses which have already been stated. In order to get measures of pre-experiment mathematics achievement, STEP Mathe-

matics, Form 3B was administered to all seventh grade participants and STEP Mathematics, Form 2B to all ninth grade participants during September, 1961. A series of five unit tests designed to measure achievement in topics specifically treated in the SMSG textbooks was administered to pupils at each level during the course of the year, as the topics were completed. The five unit tests administered to seventh grade students covered the following topics: non-metric geometry, factors and primes; rational numbers and fractions; decimals, ratio, and percent; measurement, area and volume; parallels, polygons, prisms, circle statistics and graphs. The five at the ninth grade level covered the following topics: sets, sentences, and variables; open sentences and properties of operations; real numbers, multiplication and addition of real numbers; properties of order, subtraction and division of real numbers; factors, exponents, radicals and polynomials.

At the termination of the experimental period in May, 1962, STEP Mathematics, Form 3A was administered to seventh grade students and STEP Mathematics, Form 2A to all ninth grade students. The unit tests scores and the scores on the final STEP Mathematics Tests were used as criteria of mathematics achievement.

Statistical Treatment

A combination of analysis of covariance and regression analysis was used to treat the data collected. The covariable at each grade level was the score on the pre-experiment STEP Mathematics test. It was thus possible to take into account any pre-experiment differences in mathematics achievement between the groups being compared. Criterion measures were scores on the post-experiment STEP Mathematics test and on the five unit tests administered at each grade level. As explained before, each pair of classes in each school was considered to make up a separate experimental situation or replication of the experiment. Thus there were thirty-six tests of hypotheses at the grade seven level: one for each of the six pairs of classes when the set of final STEP Mathematics scores was used as a criterion and five for each of the six pairs of classes when each of the unit test scores was used as a criterion. At the grade nine level there were forty-eight tests of the hypotheses; one for each of the eight pairs of classes when the set of final STEP Mathematics scores was used as a criterion and five for each of the eight classes when each set of the unit test scores was used as a criterion.

It was assumed that an individual student's score on any of the criterion tests was a function of his pre-experiment achievement and the particular SMSG textbook he used during the experimental period.

The model for this condition may be expressed as follows:

$$Y_a = B_aX_a + G + I_a$$
$$Y_b = B_bX_b + G + I_b$$

where:

Y_a: the criterion score of any student using the SMSG mathematics text for the college-capable

Y_b: the criterion scores of any student in the same school using the modified SMSG mathematics text

B_aX_a and B_bX_b: the regression effect of the covariable

I_a: the effect of using the SMSG mathematics text for college-capable students

I_b: the effect of using the modified SMSG mathematics text

G: error components which were assumed to be normally distributed with mean zero and variance equivalent to that of the populations from which the samples were drawn.

Normally, when analysis of covariance is used the means of the groups being compared are adjusted and an appropriate test used to determine whether the adjusted means are significantly different. However, this procedure can be used only when the slopes of the two regression lines involved are not different. In this investigation it was important that achievement as measured by the criterion tests could be studied for the entire range of the covariable. Thus the hypothesis tested was as follows:

$$H_o: \begin{cases} B_a = B_b \\ I_a = I_b \end{cases}$$ [3]

An appropriate test of this hypothesis could be used to determine simultaneously whether the slopes of the regression lines and the y-intercepts were different. These tests were based upon the calculation of an F value obtained as follows:

$$F = \frac{\dfrac{s_r^2 - s_a^2}{n_r - n_a}}{\dfrac{s_a^2}{n_a}}$$

where:

s_r^2: is the sum of squares of deviations of each observed score on the criterion for both classes being compared from a regression line computed from the combined data for the two classes.

[3] For a complete discussion of the procedure see: Henry Scheffe—*The Analysis of Variance*, John Wiley & Sons Inc., 1958 (esp. pp. 68-72).

s_a^2: the sum of square of deviations of observed criterion scores for the class using the SMSG text for the college-capable combined with the sum of squares of deviations of observed criterion scores for the class using the modified SMSG text. Each set of deviations was taken from the regression line of a particular criterion variable on the initial STEP scores for each class taken individually.

n_a: the sum of degrees of freedom for class using the SMSG text for college-capable students and the degrees of freedom for the class using the modified SMSG text.

n_r: number of degrees of freedom available if data for the two classes being compared is combined.

Wherever the hypothesis of no difference was rejected, scatter diagrams of the data involved were drawn to permit a complete analysis of the nature of the differences.

Results

Seventh Grade

Table I provides a summary of the results of the various tests of the hypothesis for the grade seven classes. The level for rejecting the hypotheses was set at 5 per cent.

TABLE I

SUMMARY OF RESULTS OF TESTS OF HYPOTHESES FOR SEVENTH GRADE CLASSES

SCHOOL	CRITERION					
	Unit Test 1	Unit Test 2	Unit Test 3	Unit Test 4	Unit Test 5	STEP 3A
1	a	a	a	a	a	a
2	r	a	a	a	a	a
3	a	a	a	a	a	a
4	a	a	a	a	a	a
5	a	a	a	a	a	a
6	r	r	r	a	r	r

a—hypothesis accepted
r—hypothesis rejected

It will be noted that five of the six rejections occurred in School 6. The scatter diagrams for these five rejections indicated that there was a tendency for the high ability students in this school who were low achievers in mathematics at the beginning of the experiment to achieve better if they used the modified SMSG mathematics text rather than the one for the college-capable students. There was also a tendency—though not so marked—for those who had been high achievers at the beginning of the experiment to achieve better on the criterion tests if they used the SMSG text for college-capable students than if they used the modified

version. This same trend was also noted for Unit Test 4 in this school though the differences were not great enough to be statistically significant. Of the remaining 29 of the cases in which the hypothesis was accepted, 14 of them showed a trend similar to that described above. In the case of Unit Test 1, for School 2 where the hypothesis of no difference was rejected, the pupils who used the modified version of the SMSG grade seven text tended to achieve better, over the entire range of the covariable, than those who used the regular text.

Ninth Grade

The results of the various tests of the hypothesis at the ninth grade level are presented in Table II. Here again the level for rejecting the hypothesis was set at 5 per cent.

TABLE II

SUMMARY OF RESULTS OF TESTS OF HYPOTHESES FOR NINTH GRADE CLASSES

SCHOOL	CRITERION					
	Unit Test 1	Unit Test 2	Unit Test 3	Unit Test 4	Unit Test 4	STEP 2A
7	a	r	a	a	a	a
8	a	r	r	r	a	a
9	a	a	a	a	a	a
10	a	r	a	a	a	a
11	a	a	a	r	r	a
12	a	a	a	a	a	a
13	a	a	r	a	r	a
14	a	r	a	r	a	a

a—hypothesis accepted
r—hypothesis rejected

When the STEP mathematics post-experiment scores were used as criterion there was no significant difference in any of the schools between the achievement of pupils using the SMSG text for the college-capable and those using the modified version of this text. Of the forty remaining tests of the hypothesis when the unit test scores were used as criteria twenty-nine were accepted and eleven were rejected. For School 7 and Unit Test 2, regression analysis revealed that students who used the modified version of the SMSG tended to achieve better over the whole range of the covariable than students who used the SMSG text for the college-capable. In School 8 there were three rejections on the basis of unit tests and in all these cases the tendency was for students who used the modified SMSG text to score higher than the students who used the SMSG text for the college capable. This tendency was most marked for those students whose pre-experiment mathematics achievement was high. For the seven remaining rejections, the same pattern as noted

for seventh grade pupils was found. That is, the lower achievers tended to score higher on the unit tests if they used the text for college-capable students. In most cases this tendency decreased for the higher achieving high ability students. However, even for the highest achieving students, using the text for the college-capable did not appear to give them any decided advantage. At this grade level in the majority of cases where the hypotheses of no difference were accepted the trend—though not great enough to be statisically significant—was for low achievers to perform better in the criterion tests if they used the modified SMSG text. This tendency, in general, decreased for higher achieving pupils.

Conclusions and Recommendations

1. The number of times the hypothesis that there was no difference between the mathematics achievement of high ability pupils who used the SMSG text for slower learners was rejected indicated an effect that could not be entirely attributed to chance.

2. In the cases where the hypothesis was rejected the most common tendency was for the lower achievers among these high ability pupils to achieve better on the criterion tests if they had used the SMSG text for slower learners. The superiority of the modified text tended in many cases to decrease for the higher achieving, high ability pupils and at the grade seven level there was some evidence to indicate that using the modified text would actually have an adverse effect upon the subsequent achievement of the most capable of these high ability pupils.

3. The variability in the nature of the differences as revealed by regression analysis would indicate that there are factors other than level of textbook difficulty which affect the mathematics achievement of high ability pupils at these levels. The factors are undoubtedly associated with the teacher, the pupils, the school, or various combinations of these.

4. Since the modified versions of the SMSG texts at both the grade seven and grade nine levels tended to facilitate the learning of mathematics on the part of high ability, lower achieving pupils involved in this experiment, there can be little doubt that modifications of the kind made by the School Mathematics Study Group would assist lower ability students in learning mathematics.

5. No attempt was made in this investigation to determine the relative effectiveness of the various devices used in the modified SMSG texts to simplify presentation. There is a need for this kind of study. There is also a need to determine which devices are most effective for simplifying the presentation of specific topics.

6. The term "high ability" in this investigation was not rigorously defined. There is a need for more information about the com-

parative performance of students within specifically-stated ranges of ability when they use the SMSG mathematics textbook for college-capable students or the modified version designed for slower learners. There is a need for more information about the comparative performance of students of various levels of ability on specific topics in mathematics presented in various ways. Indeed, one might question the need for providing a more difficult version of a mathematics textbook for college-capable students. Among the high ability students in this investigation only the highest achievers seemed to find the treatment which involved the use of the SMSG text for the college-capable advantageous from the standpoint of mathematics achievement. This advantage was slight and the number of students involved was small. It might be argued that, at the junior high school level, textbook writers should continually search for better ways of making the presentation of mathematical ideas as simple as possible. It would appear, for example, that reading difficulties should not be allowed to interfere with the acquisition of fundamental ideas in mathematics, that examples should be provided to promote pupil discovery of mathematical relationships and principles; and that a variety of significant problem situations should be provided to enable the student to make applications of the relationship and principles.

THE HUMAN FACTOR IN CRITICAL ANALYSIS

Norman Cuthbertson

I. Introduction

A prime characteristic of teaching has been its resistance to change. Most teachers, it seems, teach as they were taught. Moreover, teaching has been regarded as a rather simple task involving little more than the imparting of knowledge in neatly packaged quantities to usually receptive children. The assumption being that, if the teacher knew his subject, little else was required. Good teaching, therefore, was easily defined. Its hallmarks were faithful adherence to a prescribed curriculum and a devotion to exposition as a teaching technique. Schwab (7) describes the traditional dogmatic classroom as one in which "the role of the teacher was to explain what the book left unclear and to test the student's grasp of what he was told."

Behind the close door of her classroom the teacher taught as she chose or, as was more often the case, as best she knew how. The excitment of the sixties, in education, centers around an effort to break through these dual concepts of insularity and self-sufficiency which long have characterized the role of the teacher in America education. Such a breakthrough, it is hoped, would help to effect an improvement in the learning situation for pupils. Team teaching seeks to achieve this end by giving groups of teachers joint responsibility for the education of a large number of students. Under such a means the onus for improvement rests on the group, in that it assumes the prime responsibility for meeting the needs of all students under its care. Critical analysis, on the other hand, seeks to batter down the walls of complacent insularity which have long encompassed teaching to lay bare the facts of the matter. Its technique of analysis, based on data gathered by direct observation of the teacher in action, tends to focus directly on the individual concerned. Solely accountable for the amount of learning that occurs in his classes, the teacher is placed under much greater stress than is the case when members of a group are held jointly responsible.

II. Delineation and Potential

Let us now turn to a brief delineation of the procedure previously referred to as critical analysis and to some consideration of its potential. In the first place, it is not intended as a one-shot treatment. Rather it should be regarded as a continuing process which should not be instituted until a climate conducive to growth

is established. The pattern proposed at the Harvard-Lexington Summer Program includes a six-step supervisory cycle consisting of: (1) planning a lesson with a teacher, (2) observation of the teaching, (3) analysis of the teaching, (4) critique or analysis session with the teacher observed, (5) observation follow-up the next day, and (6) critique follow-up the next day. Ideally the cycle would be repeated as often as it was deemed necessary or as it was productive of the desired results. Planning of the lesson with the teacher makes it possible for the supervisor to become familiar with the objectives of the lesson as well as with the content and the strategies to be employed. In the second phase it is intended that the supervisor remain in the classroom activity observing the teacher throughout the progress of the lesson. Thereafter, time should be provided for both the teacher and the supervisor to collect their thoughts but, in particular, for the supervisor to analyse the exhaustive notes or transcriptions taken in an attempt to discover patterns of teacher behavior that bore directly on the success or failure of the lesson. The strategies employed by the teacher to achieve his objectives must be examined in the same light. In the ensuing face to face critique or analysis session, the supervisor attempts to present the evidence gathered in a manner conducive to self-analysis on the part of the teacher. Thus, it is hoped that by making the teacher aware of his strengths and, more particularly, of his weaknesses in the highly charged atmosphere of the analysis session, he will be strongly motivated to improve his methods. But the reactions of teachers to this kind of stress are as yet undertermined and the effects of continued exposure unknown. Be that as it may, the need to update and upgrade teaching is so great that we must be prepared to face the dangers. Indeed it may not be always possible or desirable to separate instruction from instructor, and the strength of the procedure may lie in the acceptance of this fact by both parties. If we accept Alpert's (1) suggestion that dominant motives in a motivational hierarchy are aversive or tensional in nature, we need not be too concerned about any momentarily advere effects a discreetly conducted analysis session may produce. However, since the supervisor must assume responsibility for the outcome, he must be aware of the dangers inherent in the technique.

This paper, therefore, proposes to examine the reactions of teachers at the 1963 Harvard-Lexington Summer Program to their initial exposures to critical analysis in an effort to test the hypothesis that instruction is not likely to improve unless it is subjected to continued analysis.

III. Presentation and Analysis of Data

A research specialist examining this paper and discerning a curious combination of the techniques of direct observation and use of a questionnaire for gathering the data here presented will, it is hoped, recognize the need to gather as much data as possible in the few days available to the writer. So too, will he recognize that the author has drawn heavily on his own experience, both past and present, in arriving at conclusions.

Tables I and II present in summary form the responses of teachers to the first set of items in each of parts one and two of the questionnaire.

TABLE I

In the following check the item that best describes how you felt on the first occasion that your teaching was subjected to critical analysis at the 1963 HLSP.

Extremely nervous	11%	22%
Very nervous	11%	
Somewhat nervous	29%	
A little nervous	42%	50%
Not at all nervous	8%	

TABLE II

In the following check the item that *best* describes how you now feel when your teaching is being subjected to critical analysis. (At end of second round of teaching.)

Extremely nervous	0%	8%
Very nervous	8%	
Somewhat nervous	18%	
A little nervous	42%	74%
Not at all nervous	32%	

Note from Table I than 22% of the responses lie in the upper two categories which indicate a highly nervous state on the part of the teacher whose work is being subjected to analysis. On the other hand, 50% of the responses lie in the lower two categories indicating little or no nervousness. From this we might conclude that the suggestion relative to the effects of stress made earlier in this paper is not borne out. Therefore, let us consider Table II which indicates a strong shift in favor of little or no nervousness. The corresponding figures are 8% for the upper two categories and 74% for the lower two. This shift is statistically significant. Thus, it appears that the feeling of nervouness experienced by teachers at Harvard-Lexington on their first exposure to critical analysis diminished rapidly as the technique became more familiar. It might also be pointed out that none of the respondents indicated extreme nervousness after his second or third experience with critical analysis. Even greater significance may be attached to this shift when consideration is given to the fact that the questionnaire was distributed from one to three weeks after the respondents

had their first analysis sessions, and the time lapse may have contributed to some modification in their recollection of the event.

One further item is worthy of record. Teachers have frequently commented that they were more nervous during the observation of their work than during the critique. Three respondents took the time to indicate this on the side of the questionnaire. Here, then, is an area that may prove fruitful for further research.

Tables III, IV, V and VI are grouped to make it possible to observe whether or not the trend indicated in Tables I and II is continued. Although the items from which the data are assembled were included in the questionnaire primarly as reinforcement for the first ones, they have some individual merit.

Table III

In the following check the item that *best* describes how you felt on the *first* occasion that your teaching was subjected to critical analysis at the 1963 HLSP.

Extremely apprehensive	11%	18%
Very apprehensive	7%	
Somewhat apprehensive	29%	
A little apprehensive	49%	53%
Not at all apprehensive	4	

Table IV

In the following check the item that best describes how you now feel when your teaching is being subjected to critical analysis (at end of second round of teaching).

Extremely apprehensive	0%	11%
Very apprehensive	11%	
Somewhat apprehensive	22%	
A little apprehensive	32%	67%
Not at all apprehensive	35%	

Table V

In the following check the item that best describes how you felt on the **first** occasion that your teaching was subjected to critical analysis at the 1963 HLSP.

Extremely insecure	11%	25%
Very insecure	14%	
Somewhat insecure	11%	
A little insecure	36%	65%
Not at all insecure	29%	

Table VI

In the following check the item that best describes how you felt when you teaching was being subjected to critical analysis (at end of second round of teaching).

Extremely insecure	0%	4%
Very insecure	4%	
Somewhat insecure	25%	
A little insecure	25%	71%
Not at all insecure	46%	

Excluding minor variations, the trend indicated in the first two tables is again evident. Since the median response in Tables I, III and V lies between the "somewhat" and "a little" categories, it appears that the teachers generally were not unduly nervous, apprehensive or insecure on the first occasion that they experienced critical analysis. Moreover, the shift in this median in Tables II, IV and VI to the middle of the "a little" category is statistically significant. If so favorable a shift in teacher reactions occurred in from one to three weeks, it may be assumed that a long-term program of analysis aimed at the improvement of instruction might early reach the stage where these kinds of emotional reactions are not longer a major consideration.

The shift in the two upper levels of from 25% in Table V to 4% in the Table VI is most significant. It appears that much of the insecurity felt by teachers during the initial sessions was due to uncertainty about what was going to happen. Once the teacher had experienced analysis, and had become aware of its purpose, feelings of insecurity faded quickly. This is further born out by the fact that 46% of all respondents indicated that they felt "not at all insecure" in later sessions.

Comparing Table IV with Tables II and VI, we note that the range from 11% to 67% for the spread from "extremely" and "very apprehensive" to "little" and "not at all apprehensive" is not so great as it is for the same categories dealing with feelings of nervousness and insecurity. These two range from 8% to 74% and from 4% to 71%, respectively. The question which now arises relates to the significance we can attach to this difference. A re-examination of the questionnaires reveals that at least three of the respondents felt more apprehensive during their later experiences with critical analysis, but none felt more nervous or more insecure. Two reasons for this increased feeling of apprehension appear valid. The first is that the teacher had not anticiapted the probing nature of the analysis session and was apprehensive about further revelations of weakness in his teaching. The second, and more plausible one, is that the teacher reacted unfavorably to the strategy used by the analyst and was apprehensive about the conduct of future sessions.

The evidence to date appears to negate the contention that the danger of unfavorable reaction among teachers would tend to severely restrict the use of critical analysis. On the contrary, we have evidence that most teachers at the 1963 Harvard-Lexington Summer Program adjusted readly to his new approach to the problem of the improvement of instruction.

Let us now turn to an examination of the data provided by the second half of the questionnaire.

TABLE VII

Indicate the extent that your teaching has changed while you have been at HLSP 1963 by checking the appropriate item.

No change	7%	36%
A little change	29%	
Some change	25%	
A great amount of change	25%	39%
An extreme change	14%	

Since 93% of all teacher polled indicated that their teaching had changed, and since the median response fell in the middle or "some change" category, it may safely be assumed that there were factors at work at Harvard-Lexington which contributed to this change. The next item on the questionnaire was included in an effort to assess the importance of the factor of critical analysis to this process of change. The teachers were asked to rank the following seven items in order of their influence on any change that may have occurred:

1. Discipline meetings with subject specialists.
2. Cross discipline team meetings.
3. Lectures at Estabrook.
4. Observation of analysis sessions of other teachers.
5. Critical analysis sessions of your teaching.
6. Observation of other teachers in action.
7. Informal discussions with other teachers.

Table VIII presents the tabulation of responses with respect to the influence of critical analysis only.

It will be noted that 68% of all respondents ranked critical analysis of their teaching as being among the first three influences on any changes that may have occurred in their teaching while they were students at Harvard-Lexington. Conversely only 20% ranked it lower than fourth in influence. Although critical analysis was ranked well ahead of all other influences, it is worthy of note that the opportunity to observe other teachers in action was ranked second in importance.

TABLE VIII

Rank order for critical analysis.

First	25%	68%
Second	18%	
Third	25%	
Fourth	11%	
Fifth	14%	20%
Sixth	3%	
Seventh	3%	

At the end of the questionnaire, space was reserved for any comments teachers cared to make about their personal reaction to the type of analysis experienced at Harvard-Lexington. Only 25% of the respondents chose not to comment! Of the 25% whose comments could be classed as unfavorable, about two-fifths felt that the supervisors did not show sufficient concern for the feelings of

teachers. Another two-fifths felt that the sessions were superficial and did not probe deeply enough. About 35% of all comments were favorable, with the remaining 15% vacillating between a modicum of praise and a neutral position. Thus, the statement of the observation team leader who commented that 40% success was a rather good batting average appears to be a valid assessment. However, in our jubilation over our successes we must not overlook the distress of the minority who chose to express concern over the apparent lack of consideration for the feelings of teachers.

Finally, from Table IX, we can obtain some conception of who the person might be whom teachers would be most favorably disposed to accept as a critic of their teaching.

TABLE IX

If critical analysis were to become a regular feature of the supervisory program in your school system this fall, whom would you prefer as your supervisor Check one.

29% the principal of your school.
29% a departmental head or subject specialist from your school.
3% a team leader in your school.
14% a supervisor employed by the school system.
11% an independent person from outside the system.
14% other .....

By combining the first three items we note that 61% of all respondents prefer to be supervised by someone attached to the staff of the school. Add to this the responses under "other" which favored a team of the teacher's peers and the figure reaches the two-thirds point. The small vote of only 3% for a team leader may be attributed to lack of experience, on the part of most respondents, with team teaching. Similiary, the light vote of 11% for an independent person indicates little interest due to lack of experience with this type of supervision. How, then, may one interpret the meager 14% response in favor of a supervisor employed by the school system? Certainly this adds credence to the notion that teachers generally have an unfavorable image of supervisors. Furthermore, four respondents added some such comment as "it depends on the person." It would appear, therefore, that the selection of the person to be charged with the responsibility of using the technique of critical analysis will be a most important factor in the success of a program.

IV. Summary and Conclusions

It has been frequently intimated that the writer was agreeably surprised by the fact that teachers generally did not react unfavorably to critical analysis. Certainly the evidence is sufficiently encouraging to merit continued experimentation and, as trained personnel become available, to begin its use in a more direct attack

on the problem of the improvement of instruction. But there are other considerations which must be given high priority if the technique of critical analysis is to reach its potential. First among these is the need to upgrade supervision itself in the hope of improving the image of the supervisor in the eyes of teachers. A second consideration is that of bringing teachers to an acceptance of supervision as a professional necessity. Teachers, it appears, want to improve their teaching. Therefore, it behooves the supervisor to improve his professional competencies to the point where his willingness to work with a teacher in an honest effort to seek improvement in the level of instruction is acceptable to the teacher. This approach is quite compatible with the dualism of technical and human skills which Lucio and McNeil (4) believe to be essential for intelligent activity on the part of the supervisor.

REFERENCES

1. Alpert, Richard, *The Shaping of Motives For Learning.* (Mimeographed).
2. Bartky, John A., *Supervision As Human Relations.* Boston: D. C. Heath and Company, 1953.
3. Bruner, Jerome, *The Process of Education.* Cambridge: Harvard University Press, 1960.
4. Lucio, W. H. and McNeil, J. D., *Supervision—A Synthesis of Thought and Action.* McGraw Book Company, Inc., 1962.
5. Morse, Arthur D., *Schools of Today, Tomorrow.* Garden City, New York: Doubleday and Company, Inc., 1960.
6. Reeves, Arthur and Melsness, Harold and Cheal, John, *Educational Administration*—The Role of the Teacher. Toronto: The MacMillan Company of Canada, Ltd., 1962.
7. Schwab, Joseph and Brandwein, Paul, *The Teaching of Science.* Cambridge: Harvard University Press, 1962.
8. Selltz-Jahoda-Deutsch-Cook, *Research Methods in Social Relations* (Revised). New York: Henry Holt and Company, Inc., 1959.
9. Worthen, John, *Processes in Education: Delineation, Potential and Realization.* Cambridge: Harvard University, 1963. (Mimeographed).

The Editorial Committee for the

ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

will be pleased to consider research articles from contributors outside the Province of Alberta, for publication in the Journal.

Manuscripts should be addressed to the Editor.

H. E. Smith